GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 05:49:35 ; Search time 1537 Seconds
(without alignments)
8994.036 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		400	84.2	473	6	AX008564	AX008564 Sequence
2	c	24	5.1	42	6	AX008582	AX008582 Sequence
3		22	4.6	29	6	AX008576	AX008576 Sequence
4		22	4.6	33	6	AX008574	AX008574 Sequence
5	c	22	4.6	40	6	AX008579	AX008579 Sequence
6		21	4.4	65361	2	AC091863	AC091863 Homo sapi
7		21	4.4	73218	2	AC121724	AC121724 Rattus no
8		21	4.4	92250	9	AC026693	AC026693 Homo sapi
9	c	21	4.4	122697	2	CNS08C8E	AL731874 Oryza sat
10	c	21	4.4	130732	2	AP003848	AP003848 Oryza sat
11		21	4.4	139613	2	CNS08C7T	AL731744 Oryza sat
12		21	4.4	149988	2	AC011929	AC011929 Homo sapi
13	c	21	4.4	156248	9	AC027322	AC027322 Homo sapi
14		21	4.4	162227	2	AC095224	AC095224 Rattus no
15		20	4.2	29	6	AX008577	AX008577 Sequence
16		20	4.2	29	6	AX008578	AX008578 Sequence
17	c	20	4.2	41	6	AX008575	AX008575 Sequence
18	c	20	4.2	918	4	AF271901	AF271901 Bos tauru
19		20	4.2	7595	4	BTCASK35	X14908 Bovine gene
20	c	20	4.2	39416	3	CEB0001	Z69634 Caenorhabdi
21	c	20	4.2	78448	2	AC114856	AC114856 Rattus no
22		20	4.2	99135	2	AC094561	AC094561 Rattus no
23		20	4.2	111656	2	AC010449	AC010449 Homo sapi
24	c	20	4.2	128676	2	AC126877	AC126877 Rattus no
25		20	4.2	134955	10	AL714026	AL714026 Mouse DNA
26	c	20	4.2	135619	9	AL354809	AL354809 Human DNA
27	c	20	4.2	138996	2	AC008803	AC008803 Homo sapi
28		20	4.2	144841	2	AC121995	AC121995 Mus muscu
29		20	4.2	145658	9	AC093154	AC093154 Homo sapi
30	c	20	4.2	149751	2	AC048339	AC048339 Homo sapi
31	c	20	4.2	151111	2	AC103077	AC103077 Rattus no
32	c	20	4.2	157348	2	AP002446	AP002446 Homo sapi
33	c	20	4.2	162731	2	AC109162	AC109162 Mus muscu
34		20	4.2	164429	9	AC026366	AC026366 Homo sapi
35		20	4.2	166841	2	AC114523	AC114523 Rattus no
36		20	4.2	166960	2	AC015958	AC015958 Homo sapi
37		20	4.2	172189	9	AC109518	AC109518 Homo sapi
38	c	20	4.2	176224	2	AP000898	AP000898 Homo sapi
39		20	4.2	176999	2	AC053518	AC053518 Homo sapi
40	c	20	4.2	177444	2	AC084238	AC084238 Mus muscu
41	c	20	4.2	184436	9	AP000919	AP000919 Homo sapi
42	c	20	4.2	185178	2	AP000937	AP000937 Homo sapi
43		20	4.2	187310	2	AC101870	AC101870 Mus muscu
44		20	4.2	192519	10	MMU278435	AJ278435 Mus muscu
45	c	20	4.2	195907	2	AC095185	AC095185 Rattus no

ALIGNMENTS

RESULT 1
AX008564
LOCUS AX008564 473 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO9966057.
ACCESSION AX008564
VERSION AX008564.1 GI:9996114
KEYWORDS .
SOURCE garden asparagus.
ORGANISM Asparagus officinalis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;
Asparagaceae; Asparagus.
REFERENCE 1 (bases 1 to 473)
AUTHORS Draper,J., Kenton,P. and Paul,W.
TITLE Inducible promoters

```
  JOURNAL   Patent: WO 9966057-A 1 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Asparagus officinalis"
                     /db_xref="taxon:4686"
BASE COUNT      162 a    112 c     67 g    132 t
ORIGIN

  Query Match             84.2%;  Score 400;  DB 6;  Length 473;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-209;
  Matches  400;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60

Qy         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120

Qy        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180

Qy        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240

Qy        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300

Qy        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360

Qy        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAA 400
              ||||||||||||||||||||||||||||||||||||||||
Db        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAA 400


RESULT 2
AX008582/c
LOCUS       AX008582                  42 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 19 from Patent WO9966057.
ACCESSION   AX008582
VERSION     AX008582.1  GI:9996132
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Draper,J., Kenton,P. and Paul,W.
  TITLE     Inducible promoters
  JOURNAL   Patent: WO 9966057-A 19 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT       13 a      7 c     14 g      8 t
ORIGIN

  Query Match             5.1%;  Score 24;  DB 6;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 0.092;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        317 GTCCACGAACTCGTACCTTATTCC 340
              ||||||||||||||||||||||||
Db         42 GTCCACGAACTCGTACCTTATTCC 19


RESULT 3
AX008576
LOCUS       AX008576                  29 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 13 from Patent WO9966057.
ACCESSION   AX008576
VERSION     AX008576.1  GI:9996126
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Draper,J., Kenton,P. and Paul,W.
  TITLE     Inducible promoters
  JOURNAL   Patent: WO 9966057-A 13 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT        9 a      7 c      8 g      5 t
ORIGIN

  Query Match             4.6%;  Score 22;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        110 TTCCATGTCATGAGAGAAGCAC 131
              ||||||||||||||||||||||
Db          8 TTCCATGTCATGAGAGAAGCAC 29


RESULT 4
AX008574
LOCUS       AX008574                  33 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 11 from Patent WO9966057.
ACCESSION   AX008574
VERSION     AX008574.1  GI:9996124
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 33)
  AUTHORS   Draper,J., Kenton,P. and Paul,W.
  TITLE     Inducible promoters
  JOURNAL   Patent: WO 9966057-A 11 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .33
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT        6 a     10 c      7 g     10 t
ORIGIN

  Query Match             4.6%;  Score 22;  DB 6;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 TTCTTATTGCGACCTGACTCTC 25
              ||||||||||||||||||||||
Db         12 TTCTTATTGCGACCTGACTCTC 33


RESULT 5
AX008579/c
LOCUS       AX008579                  40 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 16 from Patent WO9966057.
```

```
ACCESSION   AX008579
VERSION     AX008579.1  GI:9996129
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Draper,J., Kenton,P. and Paul,W.
  TITLE     Inducible promoters
  JOURNAL   Patent: WO 9966057-A 16 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT        4 a      6 c     14 g     16 t
ORIGIN

  Query Match             4.6%;  Score 22;  DB 6;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        451 CAAAAACAACACACAACCAATC 472
              |||||||||||||||||||||
Db         40 CAAAAACAACACACAACCAATC 19


RESULT 6
AC091863
LOCUS       AC091863               65361 bp    DNA     linear   HTG 09-JUN-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2350I5, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC091863
VERSION     AC091863.1  GI:14333799
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65361)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65361)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 762146
            Center clone name: CITB-H1_2350I5
            -------
            Summary Statistics
            Consensus quality: 54296 bases at least Q40
            Consensus quality: 56671 bases at least Q30
            Consensus quality: 57432 bases at least Q20
            Estimated insert size: 168010; agarose-fp estimation
            Estimated insert size: 64061; sum-of-contigs estimation
            Quality coverage:  5.38 in Q20 bases; agarose-fp estimation
            Quality coverage:  14.12 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1343: contig of 1343 bp in length
            *     1344     1443: gap of unknown length
            *     1444     2540: contig of 1097 bp in length
            *     2541     2640: gap of unknown length
            *     2641     3711: contig of 1071 bp in length
            *     3712     3811: gap of unknown length
            *     3812     4821: contig of 1010 bp in length
            *     4822     4921: gap of unknown length
            *     4922     6127: contig of 1206 bp in length
            *     6128     6227: gap of unknown length
            *     6228    11910: contig of 5683 bp in length
            *    11911    12010: gap of unknown length
            *    12011    14847: contig of 2837 bp in length
            *    14848    14947: gap of unknown length
            *    14948    18476: contig of 3529 bp in length
            *    18477    18576: gap of unknown length
            *    18577    25093: contig of 6517 bp in length
            *    25094    25193: gap of unknown length
            *    25194    33110: contig of 7917 bp in length
            *    33111    33210: gap of unknown length
            *    33211    39795: contig of 6585 bp in length
            *    39796    39895: gap of unknown length
            *    39896    46742: contig of 6847 bp in length
            *    46743    46842: gap of unknown length
            *    46843    57822: contig of 10980 bp in length
            *    57823    57922: gap of unknown length
            *    57923    65361: contig of 7439 bp in length.
FEATURES             Location/Qualifiers
     source          1. .65361
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2350I5"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    19445 a  13085 c  11801 g  19653 t   1377 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 2;  Length 65361;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
Db      16620 CTACATTAATTACAACTTTTG 16640


RESULT 7
AC121724
LOCUS       AC121724               73218 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-332D19, *** SEQUENCING IN PROGRESS
            ***, 39 unordered pieces.
ACCESSION   AC121724
VERSION     AC121724.2  GI:21909372
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 73218)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
```

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 73218)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 73218)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 19, 2002 this sequence version replaced gi:21039660.

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GYJC
    Center clone name: CH230-332D19
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 30431 bases at least Q40
    Consensus quality: 33043 bases at least Q30
    Consensus quality: 35017 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1698: contig of 1698 bp in length
*     1699     1798: gap of unknown length
*     1799     2976: contig of 1178 bp in length
*     2977     3076: gap of unknown length
*     3077     4713: contig of 1637 bp in length
*     4714     4813: gap of unknown length
*     4814     6354: contig of 1541 bp in length
*     6355     6454: gap of unknown length
*     6455     7735: contig of 1281 bp in length
*     7736     7835: gap of unknown length
*     7836     8905: contig of 1070 bp in length
*     8906     9005: gap of unknown length
*     9006    10661: contig of 1656 bp in length
*    10662    10761: gap of unknown length
*    10762    12131: contig of 1370 bp in length
*    12132    12231: gap of unknown length
*    12232    13749: contig of 1518 bp in length
*    13750    13849: gap of unknown length
*    13850    15819: contig of 1970 bp in length
*    15820    15919: gap of unknown length
*    15920    16935: contig of 1016 bp in length
*    16936    17035: gap of unknown length
*    17036    18637: contig of 1602 bp in length
*    18638    18737: gap of unknown length
*    18738    19971: contig of 1234 bp in length
*    19972    20071: gap of unknown length
*    20072    21378: contig of 1307 bp in length
*    21379    21478: gap of unknown length
*    21479    22528: contig of 1050 bp in length
*    22529    22628: gap of unknown length
*    22629    24018: contig of 1390 bp in length
*    24019    24118: gap of unknown length
*    24119    25622: contig of 1504 bp in length
*    25623    25722: gap of unknown length
*    25723    26949: contig of 1227 bp in length
*    26950    27049: gap of unknown length
*    27050    28440: contig of 1391 bp in length
*    28441    28540: gap of unknown length
*    28541    29872: contig of 1332 bp in length
*    29873    29972: gap of unknown length
*    29973    31074: contig of 1102 bp in length
*    31075    31174: gap of unknown length
*    31175    33423: contig of 2249 bp in length
*    33424    33523: gap of unknown length
*    33524    35399: contig of 1876 bp in length
*    35400    35499: gap of unknown length
*    35500    37239: contig of 1740 bp in length
*    37240    37339: gap of unknown length
*    37340    38482: contig of 1143 bp in length
*    38483    38582: gap of unknown length
*    38583    40363: contig of 1781 bp in length
*    40364    40463: gap of unknown length
*    40464    41882: contig of 1419 bp in length
*    41883    41982: gap of unknown length
*    41983    43509: contig of 1527 bp in length
*    43510    43609: gap of unknown length
*    43610    45690: contig of 2081 bp in length
*    45691    45790: gap of unknown length
*    45791    47012: contig of 1222 bp in length
*    47013    47112: gap of unknown length
*    47113    49636: contig of 2524 bp in length
*    49637    49736: gap of unknown length
*    49737    50991: contig of 1255 bp in length
*    50992    51091: gap of unknown length
*    51092    53431: contig of 2340 bp in length
*    53432    53531: gap of unknown length
*    53532    55478: contig of 1947 bp in length
*    55479    55578: gap of unknown length
*    55579    58230: contig of 2652 bp in length
*    58231    58330: gap of unknown length
*    58331    61697: contig of 3367 bp in length
*    61698    61797: gap of unknown length
*    61798    65855: contig of 4058 bp in length
*    65856    65955: gap of unknown length
*    65956    68849: contig of 2894 bp in length
*    68850    68949: gap of unknown length
```

```
          *    68950    73218: contig of 4269 bp in length.
FEATURES             Location/Qualifiers
     source          1. .73218
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-332D19"
BASE COUNT    20948 a  15112 c  14120 g  19150 t   3888 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 2;  Length 73218;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        250 TTACAACTTTTGCAAATAAAA 270
              |||||||||||||||||||||
Db      27829 TTACAACTTTTGCAAATAAAA 27849


RESULT 8
AC026693
LOCUS       AC026693               92250 bp    DNA     linear   PRI 03-OCT-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-261E10, complete sequence.
ACCESSION   AC026693
VERSION     AC026693.5  GI:15887318
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 92250)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92250)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 92250)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 92250)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 3, 2001 this sequence version replaced gi:10044353.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.4.
FEATURES             Location/Qualifiers
     source          1. .92250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-261E10"
BASE COUNT    28028 a  17238 c  17152 g  29832 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 9;  Length 92250;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
Db      58285 CTACATTAATTACAACTTTTG 58305


RESULT 9
CNS08C8E/c
LOCUS       CNS08C8E               122697 bp    DNA     linear   HTG 29-APR-2002
DEFINITION  Oryza sativa chromosome 12 clone OSJNBb0077C18, *** SEQUENCING IN
            PROGRESS ***, 4 ordered pieces.
ACCESSION   AL731874
VERSION     AL731874.1  GI:20372824
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 122697)
  AUTHORS   Choisne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
            Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
            Weissenbach,J. and Quetier,F.
  TITLE     Oryza sativa chromosome 12 sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 122697)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.
            Work on the sequence is in progress and the release of this data is
            based on the understanding that the sequence may change as work
            continue. The sequence may be contaminated with foreign sequence
            from E.coli, yeast, vector, phage, etc.
            Contigs composition :
            11793 bp contig from 1 to 11793
            2451 bp contig from 11894 to 14344
            11789 bp contig from 14445 to 26233
            96364 bp contig from 26334 to 122697.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    11793: contig of 11793 bp in length
            *   11794 11893: gap of      100 bp
            *   11894    14344: contig of 2451 bp in length
            *   14345 14444: gap of      100 bp
            *   14445    26233: contig of 11789 bp in length
            *   26234 26333: gap of      100 bp
            *   26334   122697: contig of 96364 bp in length.
FEATURES             Location/Qualifiers
     source          1. .122697
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="12"
                     /clone="OSJNBb0077C18"
                     /clone_lib="OSJNBb"
BASE COUNT    35100 a  27094 c  26028 g  34171 t    304 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 2;  Length 122697;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        131 CATGACTAAAGTAATTAGCTT 151
              |||||||||||||||||||||
Db     100162 CATGACTAAAGTAATTAGCTT 100142
```

```
RESULT 10
AP003848/c
LOCUS       AP003848              130732 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OJ1715_A07, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP003848
VERSION     AP003848.1  GI:14595193
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OJ1715_A07.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1715_A07
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 130732)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .130732
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="OJ1715_A07"
BASE COUNT    36776 a   28066 c   27679 g   38160 t     51 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 2;  Length 130732;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        131 CATGACTAAAGTAATTAGCTT 151
              |||||||||||||||||||||
Db      78724 CATGACTAAAGTAATTAGCTT 78704


RESULT 11
CNS08C7T
LOCUS       CNS08C7T              139613 bp    DNA     linear   HTG 15-APR-2002
DEFINITION  Oryza sativa chromosome 12 clone OSJNBa0026C14, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL731744
VERSION     AL731744.1  GI:20160279
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 139613)
  AUTHORS   Choisne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
            Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
            Weissenbach,J. and Quetier,F.
  TITLE     Oryza sativa chromosome 12 sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139613)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence.
            Work on the sequence is in progress and the release of this data is
            based on the understanding that the sequence may change as work
            continue. The sequence may be contaminated with foreign sequence
            from E.coli, yeast, vector, phage, etc.
            Contigs composition :
            3786 bp contig from 1 to 3786
            4831 bp contig from 3887 to 8717
            3718 bp contig from 8818 to 12535
            4235 bp contig from 12636 to 16870
            6706 bp contig from 16971 to 23676
            12914 bp contig from 23777 to 36690
            2950 bp contig from 36791 to 39740
            18336 bp contig from 39841 to 58176
            18030 bp contig from 58277 to 76306
            12219 bp contig from 76407 to 88625
            2501 bp contig from 88726 to 91226
            2193 bp contig from 91327 to 93519
            2004 bp contig from 93620 to 95623
            11756 bp contig from 95724 to 107479
            32034 bp contig from 107580 to 139613.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .139613
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="12"
                     /clone="OSJNBa0026C14"
                     /clone_lib="OSJNBa"
BASE COUNT    40605 a   29612 c   29656 g   38336 t    1404 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 2;  Length 139613;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        131 CATGACTAAAGTAATTAGCTT 151
              |||||||||||||||||||||
Db     136938 CATGACTAAAGTAATTAGCTT 136958


RESULT 12
AC011929
LOCUS       AC011929              149988 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens clone RP11-16B11, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC011929
VERSION     AC011929.3  GI:7107941
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149988)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-16B11
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 149988)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6453978.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3509
                Center clone name: 16_B_11
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 125296 bases at least Q40
                Consensus quality: 138924 bases at least Q30
                Consensus quality: 146180 bases at least Q20
                Insert size: 148000; agarose-fp
                Insert size: 149388; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1        2: contig of 2 bp in length
            *        3 102: gap of      100 bp
            *      103     4217: contig of 4115 bp in length
            *     4218 4317: gap of      100 bp
            *     4318    11679: contig of 7362 bp in length
            *    11680 11779: gap of      100 bp
            *    11780    26507: contig of 14728 bp in length
            *    26508 26607: gap of      100 bp
            *    26608    60280: contig of 33673 bp in length
            *    60281 60380: gap of      100 bp
            *    60381   103134: contig of 42754 bp in length
            *   103135 103234: gap of      100 bp
            *   103235   149988: contig of 46754 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149988
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-16B11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    103. .4217
                     /note="assembly_fragment"
     misc_feature    4318. .11679
                     /note="assembly_fragment"
     misc_feature    11780. .26507
                     /note="assembly_fragment"
     misc_feature    26608. .60280
                     /note="assembly_fragment"
     misc_feature    60381. .103134
                     /note="assembly_fragment"
     misc_feature    103235. .149988
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    47010 a  27490 c  27742 g  47146 t    600 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 2;  Length 149988;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
Db     141358 CTACATTAATTACAACTTTTG 141378


RESULT 13
AC027322/c
LOCUS       AC027322              156248 bp    DNA     linear   PRI 06-DEC-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2027G10, complete sequence.
ACCESSION   AC027322
VERSION     AC027322.4  GI:17386250
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156248)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156248)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 156248)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 6, 2001 this sequence version replaced gi:9256705.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source          1. .156248
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2027G10"
BASE COUNT    49224 a  29923 c  29730 g  47371 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 9;  Length 156248;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
```

```
Db      124126 CTACATTAATTACAACTTTTG 124106

RESULT 14
AC095224
LOCUS       AC095224               162227 bp    DNA     linear   HTG 12-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-9M19, *** SEQUENCING IN PROGRESS ***,
            77 unordered pieces.
ACCESSION   AC095224
VERSION     AC095224.3  GI:21728983
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 162227)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 162227)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 11, 2002 this sequence version replaced gi:17942456.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GDKB
               Center clone name: CH230-9M19
            -------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 95945 bases at least Q40
               Consensus quality: 100752 bases at least Q30
               Consensus quality: 104634 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 77 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1013: contig of 1013 bp in length
            *     1014     1113: gap of unknown length
            *     1114     2116: contig of 1003 bp in length
            *     2117     2216: gap of unknown length
            *     2217     3332: contig of 1116 bp in length
            *     3333     3432: gap of unknown length
            *     3433     4854: contig of 1422 bp in length
            *     4855     4954: gap of unknown length
            *     4955     6233: contig of 1279 bp in length
            *     6234     6333: gap of unknown length
            *     6334     7408: contig of 1075 bp in length
            *     7409     7508: gap of unknown length
            *     7509     8753: contig of 1245 bp in length
            *     8754     8853: gap of unknown length
            *     8854     9858: contig of 1005 bp in length
            *     9859     9958: gap of unknown length
            *     9959    11088: contig of 1130 bp in length
            *    11089    11188: gap of unknown length
            *    11189    12543: contig of 1355 bp in length
            *    12544    12643: gap of unknown length
            *    12644    14118: contig of 1475 bp in length
            *    14119    14218: gap of unknown length
            *    14219    15523: contig of 1305 bp in length
            *    15524    15623: gap of unknown length
            *    15624    16842: contig of 1219 bp in length
            *    16843    16942: gap of unknown length
            *    16943    18128: contig of 1186 bp in length
            *    18129    18228: gap of unknown length
            *    18229    19293: contig of 1065 bp in length
            *    19294    19393: gap of unknown length
            *    19394    20487: contig of 1094 bp in length
            *    20488    20587: gap of unknown length
            *    20588    21906: contig of 1319 bp in length
            *    21907    22006: gap of unknown length
            *    22007    23478: contig of 1472 bp in length
            *    23479    23578: gap of unknown length
            *    23579    25150: contig of 1572 bp in length
            *    25151    25250: gap of unknown length
            *    25251    27091: contig of 1841 bp in length
            *    27092    27191: gap of unknown length
            *    27192    29107: contig of 1916 bp in length
            *    29108    29207: gap of unknown length
            *    29208    30254: contig of 1047 bp in length
            *    30255    30354: gap of unknown length
            *    30355    31476: contig of 1122 bp in length
            *    31477    31576: gap of unknown length
            *    31577    32793: contig of 1217 bp in length
            *    32794    32893: gap of unknown length
```

```
          *    32894    34419: contig of 1526 bp in length
          *    34420    34519: gap of unknown length
          *    34520    35690: contig of 1171 bp in length
          *    35691    35790: gap of unknown length
          *    35791    37618: contig of 1828 bp in length
          *    37619    37718: gap of unknown length
          *    37719    39685: contig of 1967 bp in length
          *    39686    39785: gap of unknown length
          *    39786    42057: contig of 2272 bp in length
          *    42058    42157: gap of unknown length
          *    42158    43803: contig of 1646 bp in length
          *    43804    43903: gap of unknown length
          *    43904    45160: contig of 1257 bp in length
          *    45161    45260: gap of unknown length
          *    45261    46529: contig of 1269 bp in length
          *    46530    46629: gap of unknown length
          *    46630    48238: contig of 1609 bp in length
          *    48239    48338: gap of unknown length
          *    48339    49939: contig of 1601 bp in length
          *    49940    50039: gap of unknown length
          *    50040    51456: contig of 1417 bp in length
          *    51457    51556: gap of unknown length
          *    51557    53592: contig of 2036 bp in length
          *    53593    53692: gap of unknown length
          *    53693    54942: contig of 1250 bp in length
          *    54943    55042: gap of unknown length
          *    55043    56368: contig of 1326 bp in length
          *    56369    56468: gap of unknown length
          *    56469    59084: contig of 2616 bp in length
          *    59085    59184: gap of unknown length
          *    59185    60897: contig of 1713 bp in length
          *    60898    60997: gap of unknown length
          *    60998    62162: contig of 1165 bp in length
          *    62163    62262: gap of unknown length
          *    62263    64035: contig of 1773 bp in length
          *    64036    64135: gap of unknown length
          *    64136    66274: contig of 2139 bp in length
          *    66275    66374: gap of unknown length
          *    66375    67900: contig of 1526 bp in length
          *    67901    68000: gap of unknown length
          *    68001    69827: contig of 1827 bp in length
          *    69828    69927: gap of unknown length
          *    69928    71107: contig of 1180 bp in length
          *    71108    71207: gap of unknown length
          *    71208    72617: contig of 1410 bp in length
          *    72618    72717: gap of unknown length
          *    72718    73818: contig of 1101 bp in length
          *    73819    73918: gap of unknown length
          *    73919    75744: contig of 1826 bp in length
          *    75745    75844: gap of unknown length
          *    75845    77438: contig of 1594 bp in length
          *    77439    77538: gap of unknown length
          *    77539    80159: contig of 2621 bp in length
          *    80160    80259: gap of unknown length
          *    80260    82280: contig of 2021 bp in length
          *    82281    82380: gap of unknown length
          *    82381    84156: contig of 1776 bp in length
          *    84157    84256: gap of unknown length
          *    84257    86850: contig of 2594 bp in length

  Query Match             4.4%;  Score 21;  DB 2;  Length 162227;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        250 TTACAACTTTTGCAAATAAAA 270
              |||||||||||||||||||||
Db      63077 TTACAACTTTTGCAAATAAAA 63097


RESULT 15
AX008577
LOCUS       AX008577                  29 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 14 from Patent WO9966057.
ACCESSION   AX008577
VERSION     AX008577.1  GI:9996127
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 29)
  AUTHORS   Draper,J., Kenton,P. and Paul,W.
  TITLE     Inducible promoters
  JOURNAL   Patent: WO 9966057-A 14 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .29
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="primer"
BASE COUNT       10 a      6 c      6 g      7 t
ORIGIN

  Query Match             4.2%;  Score 20;  DB 6;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        226 TTGGAAACTGAATACCTACA 245
              ||||||||||||||||||||
Db         10 TTGGAAACTGAATACCTACA 29


Search completed: July 20, 2003, 06:50:59
Job time : 1541 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 05:20:09 ; Search time 189 Seconds
(without alignments)
5659.784 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	475	100.0	475	21	AAZ29510	A.officinalis thau
c	2	24	5.1	42	21	AAZ29523	Primer-4 used for
	3	22	4.6	29	21	AAZ29516	Primer-1 for ident
	4	22	4.6	33	21	AAZ29514	Primer-1 for const
c	5	22	4.6	40	21	AAZ29519	Primer-4 for ident
c	6	22	4.6	40	21	AAZ29521	Primer-2 used for
	7	20	4.2	29	21	AAZ29517	Primer-2 for ident
	8	20	4.2	29	21	AAZ29518	Primer-3 for ident
c	9	20	4.2	41	21	AAZ29515	Primer-2 for const
	10	19	4.0	44	21	AAZ29520	Primer-1 used for
	11	19	4.0	44	21	AAZ29522	Primer-3 used for
c	12	19	4.0	47	22	AAH29638	Drosophila melanog
c	13	19	4.0	507	21	AAC44332	Arabidopsis thalia
	14	19	4.0	1143	22	AAF63749	Drosophila gustato
c	15	19	4.0	1161	21	AAC36210	Arabidopsis thalia
c	16	19	4.0	1535	21	AAC41107	Arabidopsis thalia
c	17	19	4.0	14568	24	ABL32231	Human immune syste
	18	19	4.0	25911	23	ABL02604	Drosophila melanog
	19	19	4.0	25929	23	ABL02602	Drosophila melanog
	20	18	3.8	397	24	ABK62990	Rat sequence diffe
c	21	18	3.8	487	22	AAI83600	Human polynucleoti
c	22	18	3.8	529	24	ABQ24426	Oligonucleotide fo
	23	18	3.8	529	24	ABQ24427	Oligonucleotide fo
c	24	18	3.8	860	22	AAH98734	Human EST-derived
	25	18	3.8	1205	23	ABV25859	Human prostate exp
c	26	18	3.8	1492	24	ABK35003	Human cDNA encodin
c	27	18	3.8	1609	24	ABK93442	Human breast speci
	28	18	3.8	2730	23	ABL22146	Drosophila melanog
c	29	18	3.8	3215	23	ABL10202	Drosophila melanog
c	30	18	3.8	3420	23	ABL10200	Drosophila melanog
c	31	18	3.8	5998	24	ABL33635	Human immune syste
c	32	18	3.8	6973	24	ABL33778	Human immune syste
	33	17	3.6	68	16	AAT19824	Human gene signatu
c	34	17	3.6	240	20	AAX83756	Porcine circovirus
	35	17	3.6	253	24	ABN21432	Human ORFX polynuc
c	36	17	3.6	300	21	AAA00270	Human colon cancer
	37	17	3.6	378	24	ABN21444	Human ORFX polynuc
c	38	17	3.6	392	22	AAS46855	Human G protein-co
c	39	17	3.6	392	24	ABK81633	cDNA encoding nove
	40	17	3.6	398	23	ABV03703	Human prostate exp
	41	17	3.6	429	22	AAH53822	S. epidermidis ope
	42	17	3.6	447	24	ABN91340	Staphylococcus epi
	43	17	3.6	448	21	AAC69546	Human secreted pro
c	44	17	3.6	482	24	ABK38805	cDNA encoding lung
	45	17	3.6	494	22	AAS46858	Human G protein-co

ALIGNMENTS

RESULT 1
AAZ29510
ID AAZ29510 standard; DNA; 475 BP.
XX
AC AAZ29510;
XX
DT 14-MAR-2000 (first entry)
XX
DE A.officinalis thaumatin-like PR-5 related gene (AoPRT-L) promoter.
XX
KW Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; trait;
KW non-phytoxic inducing agent; Salicylic acid; SA; BTH; transformed plant;
KW systemic activation; developmental; environmental; pathogen resistance;
KW heterologous gene; disease control; sterility; fertility;
KW fruit ripening; ds.
XX
OS Asparagus officinalis.
XX
FH Key Location/Qualifiers
FT misc_feature 100..111
FT /*tag= a
FT /note= "Homologous to Potato wound induced promoter
FT sequence"
FT misc_feature 205..220
FT /*tag= b
FT /note= "Homologous to Tobacco PR-2 promoter sequence"
FT misc_signal 226..341
FT /*tag= c
FT /label= SA_responsive_element
FT /note= "Induced by non-phytoxic agents like Salicylic
FT acid or BTH"
FT misc_feature 251..264

```
FT                   /*tag=  d
FT                   /note= "Homologous to Carrot PR-3 and PR-4 promoter
FT                         sequence"
FT   repeat_region   281..318
FT                   /*tag=  e
FT                   /rpt_type= TANDEM
FT   repeat_unit     281..298
FT                   /*tag=  f
FT                   /note= "18 bp repeat"
FT   TATA_signal     409..416
FT                   /*tag=  g
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants  -
XX
PS   Claim 4; Fig 6; 67pp; English.
XX
CC   The present DNA sequence is a novel inducible promoter, derived from
CC   Asparagus officinalis thaumatin-like PR-5 related gene, AoPRT-L. The
CC   promoter sequence has similarities with other PR promoters. The
CC   promoter is responsive to low levels of an environmentally-acceptable
CC   and non-phytoxic inducing agents, like Salicylic acid or BTH. The
CC   promoters also exhibit low levels of pathogen induced systemic
CC   activation and environmentally or developmentally induced expression.
CC   The inducible promoter sequence is used to control the expression of
CC   heterologous genes in transformed plants, especially genes whose
CC   products affect a trait of the plant, such as pathogen resistance,
CC   disease control, sterility, fertility or fruit ripening.
XX
SQ   Sequence 475 BP; 163 A; 113 C; 67 G; 132 T; 0 other;

  Query Match           100.0%;  Score 475;  DB 21;  Length 475;
  Best Local Similarity  100.0%;  Pred. No. 7.7e-236;
  Matches  475;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60

Qy         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120

Qy        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180

Qy        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240

Qy        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300

Qy        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360

Qy        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTG 420

Qy        421 CCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATCATG 475
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATCATG 475


RESULT 2
AAZ29523/c
ID   AAZ29523 standard; DNA; 42 BP.
XX
AC   AAZ29523;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-4 used for construction of AoPRT-Lx3 promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   developmental; GUS construct; multimerisation; SA responsive element;
KW   systemic activation; AoPRT-Lx3 promoter; p22-JIT60; pUC19; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants  -
XX
PS   Example 12; Page 41; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-4, used for the construction of
CC   the AoPRT-Lx3 promoter. This primer is used to amplify the region from
CC   -133 bp to -247 bp of the AoPRT-L promoter, from p22-JIT60 and cloned
CC   into pUC19.
XX
SQ   Sequence 42 BP; 13 A; 7 C; 14 G; 8 T; 0 other;

  Query Match             5.1%;  Score 24;  DB 21;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 0.028;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        317 GTCCACGAACTCGTACCTTATTCC 340
              ||||||||||||||||||||||||
Db         42 GTCCACGAACTCGTACCTTATTCC 19


RESULT 3
AAZ29516
ID   AAZ29516 standard; DNA; 29 BP.
XX
AC   AAZ29516;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-1 for identification of SA responsive element in AoPRT-L promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
```

KW developmental; GUS construct; multimerisation; SA responsive element;
KW systemic activation; Inverse PCR; IPCR; ss.
XX
OS Synthetic.
XX
PN WO9966057-A2.
XX
PD 23-DEC-1999.
XX
PF 21-JUN-1999; 99WO-GB01949.
XX
PR 19-JUN-1998; 98GB-0013345.
XX
PA (BIOG-) BIOGEMMA UK LTD.
XX
PI Draper J, Kenton P, Paul W;
XX
DR WPI; 2000-106107/09.
XX
PT Novel promoters used to control the expression of heterologous genes in
PT transformed plants -
XX
PS Example 12; Page 40; 67pp; English.
XX
CC The present DNA sequence is a PCR primer-1, used for the identification
CC and multimerisation of a salicylic acid, SA/BTH responsive element in
CC the AoPRT-L promoter region. This primer is designed to regions of the
CC AoPRT-L promoter and used along with PCR primer-4 for the construction
CC of GUS fusion constructs.
XX
SQ Sequence 29 BP; 9 A; 7 C; 8 G; 5 T; 0 other;

Query Match 4.6%; Score 22; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 TTCCATGTCATGAGAGAAGCAC 131
||||||||||||||||||||||
Db 8 TTCCATGTCATGAGAGAAGCAC 29

RESULT 4
AAZ29514
ID AAZ29514 standard; DNA; 33 BP.
XX
AC AAZ29514;
XX
DT 14-MAR-2000 (first entry)
XX
DE Primer-1 for construction of AoPRT-L promoter-GUS chimeric gene.
XX
KW Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW systemic activation; developmental; pIPCR-TA; Inverse PCR; IPCR; ss.
XX
OS Synthetic.
XX
PN WO9966057-A2.
XX
PD 23-DEC-1999.
XX
PF 21-JUN-1999; 99WO-GB01949.
XX
PR 19-JUN-1998; 98GB-0013345.
XX
PA (BIOG-) BIOGEMMA UK LTD.
XX
PI Draper J, Kenton P, Paul W;
XX
DR WPI; 2000-106107/09.
XX
PT Novel promoters used to control the expression of heterologous genes in
PT transformed plants -
XX
PS Example 3; Page 32; 67pp; English.
XX
CC The present DNA sequence is a PCR primer-1, used for the isolation of
CC AoPRT-L promoter region. The promoter sequence was obtained from pIPCR-TA
CC using this primer. This primer is designed against both the 5'and 3'ends
CC of the promoter, with extensions to provide appropriate restriction
CC sites for further cloning.
XX
SQ Sequence 33 BP; 6 A; 10 C; 7 G; 10 T; 0 other;

Query Match 4.6%; Score 22; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TTCTTATTGCGACCTGACTCTC 25
||||||||||||||||||||||
Db 12 TTCTTATTGCGACCTGACTCTC 33

RESULT 5
AAZ29519/c
ID AAZ29519 standard; DNA; 40 BP.
XX
AC AAZ29519;
XX
DT 14-MAR-2000 (first entry)
XX
DE Primer-4 for identification of SA responsive element in AoPRT-L promoter.
XX
KW Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW developmental; GUS construct; multimerisation; SA responsive element;
KW systemic activation; template; p22-JIT60; Inverse PCR; IPCR; ss.
XX
OS Synthetic.
XX
PN WO9966057-A2.
XX
PD 23-DEC-1999.
XX
PF 21-JUN-1999; 99WO-GB01949.
XX
PR 19-JUN-1998; 98GB-0013345.
XX
PA (BIOG-) BIOGEMMA UK LTD.
XX
PI Draper J, Kenton P, Paul W;
XX
DR WPI; 2000-106107/09.
XX
PT Novel promoters used to control the expression of heterologous genes in
PT transformed plants -
XX
PS Example 12; Page 40; 67pp; English.
XX
CC The present DNA sequence is a PCR primer-4, used for the identification
CC and multimerisation of a salicylic acid, SA/BTH responsive element in
CC the AoPRT-L promoter region. This primer is used together with PCR
CC primers 1-3, using p22-JIT60 as template, for the construction of
CC GUS fusion constructs.
XX
SQ Sequence 40 BP; 4 A; 6 C; 14 G; 16 T; 0 other;

Query Match 4.6%; Score 22; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 451 CAAAAACAACACACAACCAATC 472
||||||||||||||||||||||
Db 40 CAAAAACAACACACAACCAATC 19

```
RESULT 6
AAZ29521/c
ID   AAZ29521 standard; DNA; 40 BP.
XX
AC   AAZ29521;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-2 used for construction of AoPRT-Lx3 promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   developmental; GUS construct; multimerisation; SA responsive element;
KW   systemic activation; AoPRT-Lx3 promoter; p22-JIT60; pUC19; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants  -
XX
PS   Example 12; Page 41; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-2, used for the construction of
CC   the AoPRT-Lx3 promoter. This primer is used to amplify the region from
CC   0 bp to -247 bp of the AoPRT-L promoter, from p22-JIT60 and cloned into
CC   pUC19.
XX
SQ   Sequence 40 BP; 4 A; 6 C; 14 G; 16 T; 0 other;

  Query Match             4.6%;  Score 22;  DB 21;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        451 CAAAAACAACACACAACCAATC 472
              |||||||||||||||||||||
Db         40 CAAAAACAACACACAACCAATC 19


RESULT 7
AAZ29517
ID   AAZ29517 standard; DNA; 29 BP.
XX
AC   AAZ29517;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-2 for identification of SA responsive element in AoPRT-L promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   developmental; GUS construct; multimerisation; SA responsive element;
KW   systemic activation; Inverse PCR; IPCR; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants  -
XX
PS   Example 12; Page 40; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-2, used for the identification
CC   and multimerisation of a salicylic acid, SA/BTH responsive element in
CC   the AoPRT-L promoter region. This primer is designed to regions of
CC   AoPRT-L promoter and used along with PCR primer-4 for the construction
CC   of GUS fusion constructs.
XX
SQ   Sequence 29 BP; 10 A; 6 C; 6 G; 7 T; 0 other;

  Query Match             4.2%;  Score 20;  DB 21;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 3.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        226 TTGGAAACTGAATACCTACA 245
              ||||||||||||||||||||
Db         10 TTGGAAACTGAATACCTACA 29


RESULT 8
AAZ29518
ID   AAZ29518 standard; DNA; 29 BP.
XX
AC   AAZ29518;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-3 for identification of SA responsive element in AoPRT-L promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   developmental; GUS construct; multimerisation; SA responsive element;
KW   systemic activation; Inverse PCR; IPCR; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants  -
XX
PS   Example 12; Page 40; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-3, used for the identification
CC   and multimerisation of a salicylic acid, SA/BTH responsive element in
CC   the AoPRT-L promoter region. This primer is designed to regions of
CC   AoPRT-L promoter and used along with PCR primer-4 for the construction
CC   of GUS fusion constructs.
```

```
XX
SQ   Sequence 29 BP; 9 A; 7 C; 6 G; 7 T; 0 other;

  Query Match             4.2%;  Score 20;  DB 21;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 3.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        341 ACAAAGGCTTAGACTTTCCA 360
              ||||||||||||||||||||
Db         10 ACAAAGGCTTAGACTTTCCA 29


RESULT 9
AAZ29515/c
ID   AAZ29515 standard; DNA; 41 BP.
XX
AC   AAZ29515;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-2 for construction of AoPRT-L promoter-GUS chimeric gene.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   systemic activation; developmental; pIPCR-TA; Inverse PCR; IPCR; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants -
XX
PS   Example 3; Page 32; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-2, used for the isolation of
CC   AoPRT-L promoter region. The promoter sequence was obtained from pIPCR-TA
CC   using this primer. This primer is designed against both the 5'and 3'ends
CC   of the promoter, with extensions to provide appropriate restriction
CC   sites for further cloning.
XX
SQ   Sequence 41 BP; 4 A; 8 C; 14 G; 15 T; 0 other;

  Query Match             4.2%;  Score 20;  DB 21;  Length 41;
  Best Local Similarity   100.0%;  Pred. No. 3.3;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        453 AAAACAACACACAACCAATC 472
              ||||||||||||||||||||
Db         41 AAAACAACACACAACCAATC 22


RESULT 10
AAZ29520
ID   AAZ29520 standard; DNA; 44 BP.
XX
AC   AAZ29520;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-1 used for construction of AoPRT-Lx3 promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   developmental; GUS construct; multimerisation; SA responsive element;
KW   systemic activation; AoPRT-Lx3 promoter; p22-JIT60; pUC19; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants -
XX
PS   Example 12; Page 41; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-1, used for the construction of
CC   the AoPRT-Lx3 promoter. This primer is used to amplify the region from
CC   0 bp to -247 bp of the AoPRT-L promoter, from p22-JIT60 and cloned into
CC   pUC19.
XX
SQ   Sequence 44 BP; 10 A; 11 C; 10 G; 13 T; 0 other;

  Query Match             4.0%;  Score 19;  DB 21;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        226 TTGGAAACTGAATACCTAC 244
              |||||||||||||||||||
Db         26 TTGGAAACTGAATACCTAC 44


RESULT 11
AAZ29522
ID   AAZ29522 standard; DNA; 44 BP.
XX
AC   AAZ29522;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Primer-3 used for construction of AoPRT-Lx3 promoter.
XX
KW   Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; primer;
KW   non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KW   developmental; GUS construct; multimerisation; SA responsive element;
KW   systemic activation; AoPRT-Lx3 promoter; p22-JIT60; pUC19; ss.
XX
OS   Synthetic.
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
```

```
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants -
XX
PS   Example 12; Page 41; 67pp; English.
XX
CC   The present DNA sequence is a PCR primer-3, used for the construction of
CC   the AoPRT-Lx3 promoter. This primer is used to amplify the region from
CC   -133 bp to -247 bp of the AoPRT-L promoter, from p22-JIT60 and cloned
CC   into pUC19.
XX
SQ   Sequence 44 BP; 10 A; 11 C; 10 G; 13 T; 0 other;

  Query Match             4.0%;  Score 19;  DB 21;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        226 TTGGAAACTGAATACCTAC 244
              |||||||||||||||||||
Db         26 TTGGAAACTGAATACCTAC 44


RESULT 12
AAH29638/c
ID   AAH29638 standard; DNA; 47 BP.
XX
AC   AAH29638;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Drosophila melanogaster essential gene fragment, SEQ ID NO: 827.
XX
KW   Drosophila melanogaster; fruit fly; essential gene; screening assay;
KW   pesticide; crop protection; chromosome 2; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200118547-A1.
XX
PD   15-MAR-2001.
XX
PF   06-SEP-2000; 2000WO-GB03444.
XX
PR   07-SEP-1999;   99GB-0021009.
XX
PA   (UNIU ) UNIV GLASGOW.
XX
PI   Davies RW,  Kaiser K,  Yang MY;
XX
DR   WPI; 2001-281436/29.
XX
PT   Screening assays for used for identifying compounds having a
PT   physiological effect on proteins identified as being essential -
XX
PS   Claim 1; Page 642; 695pp; English.
XX
CC   The present sequence is part of an essential gene from Drosophila
CC   melanogaster. Lack of expression of the protein encoded by this
CC   gene leads to a lethal or semi-lethal phenotype. The invention
CC   relates to 902 nucleic acid sequences from genes encoding proteins
CC   which are thought to be essential, and to a screening assay for
CC   identifying compounds which have a physiological effect on these
CC   proteins. Suitable compounds are useful as pesticides and may be used
CC   in conjunction with other pesticides and herbicides for crop
CC   protection. The gene corresponding to the present sequence is located
CC   on chromosome 2.
XX
SQ   Sequence 47 BP; 13 A; 6 C; 9 G; 19 T; 0 other;

  Query Match             4.0%;  Score 19;  DB 22;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        249 ATTACAACTTTTGCAAATA 267
              |||||||||||||||||||
Db         41 ATTACAACTTTTGCAAATA 23


RESULT 13
AAC44332/c
ID   AAC44332 standard; DNA; 507 BP.
XX
AC   AAC44332;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 42470.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
```

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.0%; Score 19; DB 21; Length 507;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         11 TGCGACCTGACTCTCTTGT 29
              |||||||||||||||||||
Db        289 TGCGACCTGACTCTCTTGT 271
```

RESULT 14
AAF63749

ID AAF63749 standard; DNA; 1143 BP.
XX
AC AAF63749;
XX
DT 03-APR-2001 (first entry)
XX
DE Drosophila gustatory receptor GR39D.2c DNA sequence.
XX
KW Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
KW crop damage; pest control; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200077208-A2.
XX
PD 21-DEC-2000.
XX
PF 14-JUN-2000; 2000WO-US16211.
XX
PR 14-JUN-1999; 99US-0138668.
PR 10-FEB-2000; 2000US-0181704.
XX
PA (UYYA) UNIV YALE.
XX
PI Carlson PJ, Clyne PJ, Warr CG;
XX
DR WPI; 2001-061873/07.
DR P-PSDB; AAB75210.
XX
PT New isolated nucleic acid molecule encoding Drosophila Gustatory
PT Receptor protein useful for e.g. identification of compounds which may
PT be used for pest management -
XX
PS Claim 3; Page 134-135; 227pp; English.
XX
CC This invention relates to polynucleotide sequences AAF63732 - AAF63777
CC which encode Drosophila gustatory receptor proteins represented by
CC sequences AAB75193 - AAB75238. The invention includes methods for
CC determining gustatory receptor ligands. Also included is a method for
CC modulating the expression of the DNA encoding the receptors. The DNA and
CC protein sequences may be used for the identification of compounds,
CC e.g. pheromones and other semiochemicals, which may be used for pest
CC management. The DNA sequences may also be used for behavioural studies
CC involving gustatory systems in various organisms. Also, the DNA sequences
CC may also be used to track down gustatory receptor genes in insects that
CC damage crops or transmit diseases.
XX
SQ Sequence 1143 BP; 326 A; 232 C; 226 G; 359 T; 0 other;

Query Match 4.0%; Score 19; DB 22; Length 1143;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 249 ATTACAACTTTTGCAAATA 267
||||||||||||||||||||
Db 641 ATTACAACTTTTGCAAATA 659

RESULT 15
AAC36210/c
ID AAC36210 standard; DNA; 1161 BP.
XX
AC AAC36210;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12948.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.0%; Score 19; DB 21; Length 1161;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TGCGACCTGACTCTCTTGT 29
||||||||||||||||||||
Db 370 TGCGACCTGACTCTCTTGT 352

Search completed: July 20, 2003, 06:25:08
Job time : 190 secs

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 06:21:55 ; Search time 43 Seconds
(without alignments)
3387.708 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17	3.6	447	4	US-09-134-001C-803	Sequence 803, App
c	2	17	3.6	1767	4	US-09-347-594-1	Sequence 1, Appli
c	3	17	3.6	1767	4	US-09-347-594-2	Sequence 2, Appli
c	4	17	3.6	1767	4	US-09-082-558-1	Sequence 1, Appli
c	5	17	3.6	1767	4	US-09-082-558-2	Sequence 2, Appli
c	6	17	3.6	1767	4	US-09-161-092-1	Sequence 1, Appli
c	7	17	3.6	1767	4	US-09-161-092-2	Sequence 2, Appli
c	8	17	3.6	1768	4	US-09-347-594-3	Sequence 3, Appli
c	9	17	3.6	1768	4	US-09-347-594-4	Sequence 4, Appli
c	10	17	3.6	1768	4	US-09-082-558-3	Sequence 3, Appli
c	11	17	3.6	1768	4	US-09-082-558-4	Sequence 4, Appli
c	12	17	3.6	1768	4	US-09-082-558-6	Sequence 6, Appli
c	13	17	3.6	1768	4	US-09-161-092-3	Sequence 3, Appli
c	14	17	3.6	1768	4	US-09-161-092-4	Sequence 4, Appli
c	15	17	3.6	1768	4	US-09-161-092-6	Sequence 6, Appli
	16	17	3.6	2281	1	US-08-164-614A-1	Sequence 1, Appli
	17	17	3.6	2281	2	US-08-456-489B-1	Sequence 1, Appli
	18	17	3.6	2281	5	PCT-US93-01720-1	Sequence 1, Appli
	19	17	3.6	3935	4	US-09-455-777-1	Sequence 1, Appli
c	20	17	3.6	9048	3	US-08-973-273-4	Sequence 4, Appli
	21	17	3.6	55827	4	US-09-813-133A-3	Sequence 3, Appli
	22	16	3.4	431	4	US-09-020-956-169	Sequence 169, App
	23	16	3.4	431	4	US-09-030-607-169	Sequence 169, App
	24	16	3.4	431	4	US-09-605-785-169	Sequence 169, App
	25	16	3.4	431	4	US-09-439-313-169	Sequence 169, App
	26	16	3.4	431	4	US-09-352-616A-169	Sequence 169, App
	27	16	3.4	431	4	US-09-232-149A-169	Sequence 169, App
	28	16	3.4	522	4	US-09-134-001C-1041	Sequence 1041, Ap
	29	16	3.4	537	4	US-09-308-003-7	Sequence 7, Appli
c	30	16	3.4	720	4	US-09-540-824-6	Sequence 6, Appli
c	31	16	3.4	1620	2	US-08-874-186-37	Sequence 37, Appl
c	32	16	3.4	1716	1	US-08-348-792-1	Sequence 1, Appli
c	33	16	3.4	1716	2	US-08-462-738-1	Sequence 1, Appli
c	34	16	3.4	1716	4	US-09-199-955-1	Sequence 1, Appli
c	35	16	3.4	1716	4	US-08-880-875-1	Sequence 1, Appli
c	36	16	3.4	1761	4	US-09-134-001C-1619	Sequence 1619, Ap
c	37	16	3.4	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	38	16	3.4	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	39	16	3.4	246240	2	US-08-724-394A-22	Sequence 22, Appl
	40	15	3.2	51	1	US-07-977-696C-63	Sequence 63, Appl
	41	15	3.2	51	1	US-08-129-930B-63	Sequence 63, Appl
	42	15	3.2	51	4	US-08-976-288A-63	Sequence 63, Appl
c	43	15	3.2	112	2	US-08-449-287-23	Sequence 23, Appl
c	44	15	3.2	218	3	US-09-035-648-14	Sequence 14, Appl
c	45	15	3.2	218	4	US-09-001-951-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-803
; Sequence 803, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 803
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-803

Query Match 3.6%; Score 17; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        264 AATAAAATATAAAGAAA 280
              |||||||||||||||||
Db        255 AATAAAATATAAAGAAA 271
```

RESULT 2
US-09-347-594-1/c
; Sequence 1, Application US/09347594
; Patent No. 6217883
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/347,594
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: 98 08777
; EARLIER FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

```
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-347-594-1

  Query Match              3.6%;  Score 17;  DB 4;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 3
US-09-347-594-2/c
; Sequence 2, Application US/09347594
; Patent No. 6217883
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon M.
;  APPLICANT: MEEHAN, Brian M.
;  APPLICANT: ELLIS, John A.
;  APPLICANT: KRAKOWKA, George S.
;  APPLICANT: AUDONNET, Jean-Christophe F.
;  TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
;  FILE REFERENCE: 454313-2338
;  CURRENT APPLICATION NUMBER: US/09/347,594
;  CURRENT FILING DATE: 1999-07-01
;  EARLIER APPLICATION NUMBER: 98 08777
;  EARLIER FILING DATE: 1998-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-347-594-2

  Query Match              3.6%;  Score 17;  DB 4;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 4
US-09-082-558-1/c
; Sequence 1, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT: MEEHAN, Brian
;  APPLICANT: CLARK, Edward
;  APPLICANT: HAINES, Deborah
;  APPLICANT: HASSARD, Lori
;  APPLICANT: HARDING, John
;  APPLICANT: CHARREYRE, Catherine E.
;  APPLICANT: CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION: REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/082,558A
;  CURRENT FILING DATE: 1998-05-21
;  EARLIER APPLICATION NUMBER: FR 9800873
;  EARLIER FILING DATE: 1998-01-22
;  EARLIER APPLICATION NUMBER: FR 9803707
;  EARLIER FILING DATE: 1998-03-20
;  EARLIER APPLICATION NUMBER: FR 97/12382
;  EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-082-558-1

  Query Match              3.6%;  Score 17;  DB 4;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 5
US-09-082-558-2/c
; Sequence 2, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT: MEEHAN, Brian
;  APPLICANT: CLARK, Edward
;  APPLICANT: HAINES, Deborah
;  APPLICANT: HASSARD, Lori
;  APPLICANT: HARDING, John
;  APPLICANT: CHARREYRE, Catherine E.
;  APPLICANT: CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION: REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/082,558A
;  CURRENT FILING DATE: 1998-05-21
;  EARLIER APPLICATION NUMBER: FR 9800873
;  EARLIER FILING DATE: 1998-01-22
;  EARLIER APPLICATION NUMBER: FR 9803707
;  EARLIER FILING DATE: 1998-03-20
;  EARLIER APPLICATION NUMBER: FR 97/12382
;  EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-082-558-2

  Query Match              3.6%;  Score 17;  DB 4;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 6
US-09-161-092-1/c
; Sequence 1, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT:  MEEHAN, Brian
;  APPLICANT:  CLARK, Edward
;  APPLICANT:  HAINES, Deborah
;  APPLICANT:  HASSARD, Lori
;  APPLICANT:  HARDING, John
;  APPLICANT:  CHARREYRE, Catherine E.
;  APPLICANT:  CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION:  REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/161,092
```

```
;  CURRENT FILING DATE:  1998-09-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/082,558
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-161-092-1

  Query Match             3.6%;  Score 17;  DB 4;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 7
US-09-161-092-2/c
; Sequence 2, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT:  MEEHAN, Brian
;  APPLICANT:  CLARK, Edward
;  APPLICANT:  HAINES, Deborah
;  APPLICANT:  HASSARD, Lori
;  APPLICANT:  HARDING, John
;  APPLICANT:  CHARREYRE, Catherine E.
;  APPLICANT:  CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION:  REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/161,092
;  CURRENT FILING DATE:  1998-09-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/082,558
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1767
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-161-092-2

  Query Match             3.6%;  Score 17;  DB 4;  Length 1767;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 8
US-09-347-594-3/c
; Sequence 3, Application US/09347594
; Patent No. 6217883
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon M.
;  APPLICANT: MEEHAN, Brian M.
;  APPLICANT: ELLIS, John A.
;  APPLICANT: KRAKOWKA, George S.
;  APPLICANT: AUDONNET, Jean-Christophe F.
;  TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
;  FILE REFERENCE: 454313-2338
;  CURRENT APPLICATION NUMBER: US/09/347,594
;  CURRENT FILING DATE: 1999-07-01
;  EARLIER APPLICATION NUMBER: 98 08777
;  EARLIER FILING DATE: 1998-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-347-594-3

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 9
US-09-347-594-4/c
; Sequence 4, Application US/09347594
; Patent No. 6217883
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon M.
;  APPLICANT: MEEHAN, Brian M.
;  APPLICANT: ELLIS, John A.
;  APPLICANT: KRAKOWKA, George S.
;  APPLICANT: AUDONNET, Jean-Christophe F.
;  TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
;  FILE REFERENCE: 454313-2338
;  CURRENT APPLICATION NUMBER: US/09/347,594
;  CURRENT FILING DATE: 1999-07-01
;  EARLIER APPLICATION NUMBER: 98 08777
;  EARLIER FILING DATE: 1998-07-06
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-347-594-4

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 10
US-09-082-558-3/c
; Sequence 3, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT: MEEHAN, Brian
;  APPLICANT: CLARK, Edward
;  APPLICANT: HAINES, Deborah
;  APPLICANT: HASSARD, Lori
;  APPLICANT: HARDING, John
```

```
;  APPLICANT: CHARREYRE, Catherine E.
;  APPLICANT: CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION: REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/082,558A
;  CURRENT FILING DATE: 1998-05-21
;  EARLIER APPLICATION NUMBER: FR 9800873
;  EARLIER FILING DATE: 1998-01-22
;  EARLIER APPLICATION NUMBER: FR 9803707
;  EARLIER FILING DATE: 1998-03-20
;  EARLIER APPLICATION NUMBER: FR 97/12382
;  EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-082-558-3

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 11
US-09-082-558-4/c
; Sequence 4, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT: MEEHAN, Brian
;  APPLICANT: CLARK, Edward
;  APPLICANT: HAINES, Deborah
;  APPLICANT: HASSARD, Lori
;  APPLICANT: HARDING, John
;  APPLICANT: CHARREYRE, Catherine E.
;  APPLICANT: CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION: REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/082,558A
;  CURRENT FILING DATE: 1998-05-21
;  EARLIER APPLICATION NUMBER: FR 9800873
;  EARLIER FILING DATE: 1998-01-22
;  EARLIER APPLICATION NUMBER: FR 9803707
;  EARLIER FILING DATE: 1998-03-20
;  EARLIER APPLICATION NUMBER: FR 97/12382
;  EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-082-558-4

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 12
US-09-082-558-6/c
; Sequence 6, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT: MEEHAN, Brian
;  APPLICANT: CLARK, Edward
;  APPLICANT: HAINES, Deborah
;  APPLICANT: HASSARD, Lori
;  APPLICANT: HARDING, John
;  APPLICANT: CHARREYRE, Catherine E.
;  APPLICANT: CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION: REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/082,558A
;  CURRENT FILING DATE: 1998-05-21
;  EARLIER APPLICATION NUMBER: FR 9800873
;  EARLIER FILING DATE: 1998-01-22
;  EARLIER APPLICATION NUMBER: FR 9803707
;  EARLIER FILING DATE: 1998-03-20
;  EARLIER APPLICATION NUMBER: FR 97/12382
;  EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)..(1768)
;   OTHER INFORMATION: N represents A or C or G or T
US-09-082-558-6

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        386 GTGCTGCCGAGGTGCTG 370


RESULT 13
US-09-161-092-3/c
; Sequence 3, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT:  MEEHAN, Brian
;  APPLICANT:  CLARK, Edward
;  APPLICANT:  HAINES, Deborah
;  APPLICANT:  HASSARD, Lori
;  APPLICANT:  HARDING, John
;  APPLICANT:  CHARREYRE, Catherine E.
;  APPLICANT:  CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION:  REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/161,092
;  CURRENT FILING DATE:  1998-09-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/082,558
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1768
```

```
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-161-092-3

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 14
US-09-161-092-4/c
; Sequence 4, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT:  MEEHAN, Brian
;  APPLICANT:  CLARK, Edward
;  APPLICANT:  HAINES, Deborah
;  APPLICANT:  HASSARD, Lori
;  APPLICANT:  HARDING, John
;  APPLICANT:  CHARREYRE, Catherine E.
;  APPLICANT:  CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION:  REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/161,092
;  CURRENT FILING DATE:  1998-09-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/082,558
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
US-09-161-092-4

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        385 GTGCTGCCGAGGTGCTG 369


RESULT 15
US-09-161-092-6/c
; Sequence 6, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
;  APPLICANT: ALLAN, Gordon
;  APPLICANT:  MEEHAN, Brian
;  APPLICANT:  CLARK, Edward
;  APPLICANT:  HAINES, Deborah
;  APPLICANT:  HASSARD, Lori
;  APPLICANT:  HARDING, John
;  APPLICANT:  CHARREYRE, Catherine E.
;  APPLICANT:  CHAPPUIS, Gilles E.
;  TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
;  TITLE OF INVENTION:  REAGENTS
;  FILE REFERENCE: ALLAN
;  CURRENT APPLICATION NUMBER: US/09/161,092
;  CURRENT FILING DATE:  1998-09-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/082,558
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 1768
;   TYPE: DNA
;   ORGANISM: Porcine circovirus
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)..(1768)
;   OTHER INFORMATION: N represents A or C or G or T
US-09-161-092-6

  Query Match             3.6%;  Score 17;  DB 4;  Length 1768;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         31 GTGCTGCCGAGGTGCTG 47
              |||||||||||||||||
Db        386 GTGCTGCCGAGGTGCTG 370


Search completed: July 20, 2003, 07:21:09
Job time : 45 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 06:51:05 ; Search time 172 Seconds
(without alignments)
5697.240 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1439767 seqs, 1031500376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	4.4	716	15	US-10-027-632-34117	Sequence 34117, A
	2	21	4.4	716	15	US-10-027-632-34118	Sequence 34118, A
	3	21	4.4	1691139	15	US-10-067-514-1	Sequence 1, Appli
c	4	19	4.0	584	15	US-10-027-632-243175	Sequence 243175,
	5	19	4.0	649	15	US-10-027-632-134120	Sequence 134120,
	6	19	4.0	649	15	US-10-027-632-134121	Sequence 134121,
	7	18	3.8	397	11	US-09-917-800A-897	Sequence 897, App
c	8	18	3.8	475	12	US-09-918-995-25871	Sequence 25871, A
	9	18	3.8	491	15	US-10-027-632-48705	Sequence 48705, A
	10	18	3.8	491	15	US-10-027-632-48706	Sequence 48706, A
	11	18	3.8	612	15	US-10-027-632-61505	Sequence 61505, A
	12	18	3.8	612	15	US-10-027-632-62397	Sequence 62397, A
	13	18	3.8	612	15	US-10-027-632-63229	Sequence 63229, A
c	14	18	3.8	656	15	US-10-027-632-111366	Sequence 111366,
	15	18	3.8	694	15	US-10-027-632-48227	Sequence 48227, A
	16	18	3.8	745	15	US-10-027-632-28109	Sequence 28109, A
	17	18	3.8	745	15	US-10-027-632-139025	Sequence 139025,
c	18	18	3.8	1492	10	US-09-822-849A-141	Sequence 141, App
c	19	18	3.8	1609	15	US-10-082-830-48	Sequence 48, Appl
c	20	18	3.8	148567	11	US-09-801-876B-3	Sequence 3, Appli
c	21	18	3.8	148567	15	US-10-254-869-3	Sequence 3, Appli
c	22	18	3.8	1691139	15	US-10-067-514-1	Sequence 1, Appli
c	23	17	3.6	240	11	US-09-935-428A-12	Sequence 12, Appl
c	24	17	3.6	392	10	US-09-811-284-37	Sequence 37, Appl
c	25	17	3.6	399	12	US-09-918-995-36916	Sequence 36916, A
c	26	17	3.6	410	11	US-09-960-352-3865	Sequence 3865, Ap
	27	17	3.6	412	11	US-09-998-598-1114	Sequence 1114, Ap
c	28	17	3.6	445	12	US-09-918-995-1630	Sequence 1630, Ap
	29	17	3.6	448	15	US-10-050-704-45	Sequence 45, Appl
c	30	17	3.6	452	15	US-10-027-632-2244	Sequence 2244, Ap
c	31	17	3.6	454	15	US-10-027-632-108770	Sequence 108770,
	32	17	3.6	466	11	US-09-998-598-1522	Sequence 1522, Ap
c	33	17	3.6	482	11	US-09-736-457-843	Sequence 843, App
c	34	17	3.6	482	11	US-09-902-941-843	Sequence 843, App
c	35	17	3.6	482	11	US-09-849-626-843	Sequence 843, App
c	36	17	3.6	482	15	US-10-017-754-843	Sequence 843, App
	37	17	3.6	489	12	US-09-918-995-33867	Sequence 33867, A
	38	17	3.6	494	10	US-09-811-284-40	Sequence 40, Appl
	39	17	3.6	501	15	US-10-102-524-609	Sequence 609, App
	40	17	3.6	506	11	US-09-764-847-463	Sequence 463, App
	41	17	3.6	506	15	US-10-092-154-463	Sequence 463, App
c	42	17	3.6	541	15	US-10-027-632-321125	Sequence 321125,
	43	17	3.6	581	10	US-09-799-777-132	Sequence 132, App
c	44	17	3.6	634	15	US-10-027-632-252063	Sequence 252063,
c	45	17	3.6	645	15	US-10-027-632-274868	Sequence 274868,

ALIGNMENTS

RESULT 1
US-10-027-632-34117
; Sequence 34117, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34117
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34117

Query Match 4.4%; Score 21; DB 15; Length 716;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 CTACATTAATTACAACTTTTG 261
||||||||||||||||||||
Db 361 CTACATTAATTACAACTTTTG 381

```
RESULT 2
US-10-027-632-34118
; Sequence 34118, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34118
;   LENGTH: 716
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-34118

  Query Match             4.4%;  Score 21;  DB 15;  Length 716;
  Best Local Similarity   100.0%;  Pred. No. 0.86;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
Db        361 CTACATTAATTACAACTTTTG 381


RESULT 3
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  TITLE OF INVENTION: HUMAN STROKE GENE
;  FILE REFERENCE: 2345.2010-003
;  CURRENT APPLICATION NUMBER: US/10/067,514
;  CURRENT FILING DATE:  2002-02-04
;  PRIOR APPLICATION NUMBER: US 09/811/352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Human
US-10-067-514-1

  Query Match             4.4%;  Score 21;  DB 15;  Length 1691139;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
Db     761375 CTACATTAATTACAACTTTTG 761395


RESULT 4
US-10-027-632-243175/c
; Sequence 243175, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 243175
;   LENGTH: 584
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(584)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-243175

  Query Match             4.0%;  Score 19;  DB 15;  Length 584;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        254 AACTTTTGCAAATAAAATA 272
              |||||||||||||||||||
Db        156 AACTTTTGCAAATAAAATA 138


RESULT 5
US-10-027-632-134120
; Sequence 134120, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134120
```

```
;   LENGTH: 649
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-134120

  Query Match             4.0%;  Score 19;  DB 15;  Length 649;
  Best Local Similarity   100.0%;  Pred. No. 9.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        418 CTGCCCATTCCCCTCCTCC 436
              |||||||||||||||||||
Db        433 CTGCCCATTCCCCTCCTCC 451


RESULT 6
US-10-027-632-134121
; Sequence 134121, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134121
;   LENGTH: 649
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-134121

  Query Match             4.0%;  Score 19;  DB 15;  Length 649;
  Best Local Similarity   100.0%;  Pred. No. 9.8;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        418 CTGCCCATTCCCCTCCTCC 436
              |||||||||||||||||||
Db        433 CTGCCCATTCCCCTCCTCC 451


RESULT 7
US-09-917-800A-897
; Sequence 897, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 897
;   LENGTH: 397
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI171319
US-09-917-800A-897

  Query Match             3.8%;  Score 18;  DB 11;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        451 CAAAAACAACACACAACC 468
              ||||||||||||||||||
Db         55 CAAAAACAACACACAACC 72


RESULT 8
US-09-918-995-25871/c
; Sequence 25871, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;  TITLE OF INVENTION:  FROM VARIOUS cDNA LIBRARIES
;  FILE REFERENCE: 20411-756
;  CURRENT APPLICATION NUMBER: US/09/918,995
;  CURRENT FILING DATE:  2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 25871
;   LENGTH: 475
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(475)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-25871

  Query Match             3.8%;  Score 18;  DB 12;  Length 475;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAATAAAATATAAAGAAA 280
              ||||||||||||||||||
Db         68 AAATAAAATATAAAGAAA 51


RESULT 9
US-10-027-632-48705
; Sequence 48705, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
```

```
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48705
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-48705

  Query Match             3.8%;  Score 18;  DB 15;  Length 491;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        100 CAATAATTATTTCCATGT 117
              ||||||||||||||||||
Db        439 CAATAATTATTTCCATGT 456


RESULT 10
US-10-027-632-48706
; Sequence 48706, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48706
;   LENGTH: 491
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-48706

  Query Match             3.8%;  Score 18;  DB 15;  Length 491;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        100 CAATAATTATTTCCATGT 117
              ||||||||||||||||||
Db        439 CAATAATTATTTCCATGT 456


RESULT 11
US-10-027-632-61505
; Sequence 61505, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 61505
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-61505

  Query Match             3.8%;  Score 18;  DB 15;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAATAAAATATAAAGAAA 280
              ||||||||||||||||||
Db        402 AAATAAAATATAAAGAAA 419


RESULT 12
US-10-027-632-62397
; Sequence 62397, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62397
;   LENGTH: 612
```

```
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-62397

  Query Match             3.8%;  Score 18;  DB 15;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAATAAAATATAAAGAAA 280
              ||||||||||||||||||
Db        402 AAATAAAATATAAAGAAA 419


RESULT 13
US-10-027-632-63229
; Sequence 63229, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 63229
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-63229

  Query Match             3.8%;  Score 18;  DB 15;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAATAAAATATAAAGAAA 280
              ||||||||||||||||||
Db        402 AAATAAAATATAAAGAAA 419


RESULT 14
US-10-027-632-111366/c
; Sequence 111366, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 111366
;   LENGTH: 656
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-111366

  Query Match             3.8%;  Score 18;  DB 15;  Length 656;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        444 TCTAACTCAAAAACAACA 461
              ||||||||||||||||||
Db        385 TCTAACTCAAAAACAACA 368


RESULT 15
US-10-027-632-48227
; Sequence 48227, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48227
;   LENGTH: 694
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-48227

  Query Match             3.8%;  Score 18;  DB 15;  Length 694;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 AAATAAAATATAAAGAAA 280
              ||||||||||||||||||
Db        392 AAATAAAATATAAAGAAA 409


Search completed: July 20, 2003, 07:53:40
Job time : 176 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 06:08:35 ; Search time 1127 Seconds
(without alignments)
6825.959 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	4.4	252	10	BE370506	BE370506 601218294
	2	21	4.4	426	10	BE147853	BE147853 RC3-HT023
c	3	21	4.4	587	10	BE205326	BE205326 EST398002
c	4	21	4.4	669	13	BI311957	BI311957 EST531370
	5	21	4.4	769	17	AG143540	AG143540 Pan trogl
	6	21	4.4	822	17	BH371846	BH371846 AG-ND-162
	7	21	4.4	845	12	BF573809	BF573809 602132006
	8	21	4.4	970	17	CNS030N2	AL222455 Tetraodon
	9	20	4.2	401	10	BB818014	BB818014 BB818014
c	10	20	4.2	401	13	BI790095	BI790095 ic93b12.x
	11	20	4.2	405	9	AA461949	AA461949 vg71c09.r
	12	20	4.2	408	10	BB810434	BB810434 BB810434
c	13	20	4.2	413	17	AZ327901	AZ327901 1M0051K05
c	14	20	4.2	418	12	BF225508	BF225508 uy41f07.x
	15	20	4.2	420	13	BG975401	BG975401 602843052
c	16	20	4.2	430	12	BG143281	BG143281 ia96a04.x
c	17	20	4.2	432	9	AU043619	AU043619 AU043619
	18	20	4.2	449	13	BM502635	BM502635 ii36f08.y
c	19	20	4.2	468	12	BG071544	BG071544 H3100C04-
	20	20	4.2	472	9	AI227421	AI227421 uf09e04.y
	21	20	4.2	475	9	AA615561	AA615561 vo79c08.r
	22	20	4.2	475	10	BE448225	BE448225 ut81a07.y
c	23	20	4.2	495	10	AW214359	AW214359 uo47b03.x
c	24	20	4.2	496	17	AZ767485	AZ767485 1M0566L10
c	25	20	4.2	512	10	AW214512	AW214512 uo50a12.x
c	26	20	4.2	518	13	BM502321	BM502321 ii36f08.x
c	27	20	4.2	532	9	AU019722	AU019722 AU019722
c	28	20	4.2	538	10	BE631935	BE631935 uu09f05.x
c	29	20	4.2	542	13	BM222085	BM222085 K0114F03-
c	30	20	4.2	547	9	AU040283	AU040283 AU040283
c	31	20	4.2	552	13	BM229851	BM229851 K0286A03-
c	32	20	4.2	569	9	AU021058	AU021058 AU021058
c	33	20	4.2	572	14	BM900102	BM900102 UI-M-DJ1-
c	34	20	4.2	574	13	BM225839	BM225839 K0218A09-
c	35	20	4.2	580	13	BM228549	BM228549 K0263C06-
c	36	20	4.2	586	13	BM229869	BM229869 K0286C12-
	37	20	4.2	607	12	BG084386	BG084386 H3100C04-
c	38	20	4.2	656	13	BM228341	BM228341 K0259E03-
	39	20	4.2	657	17	AZ055055	AZ055055 RPCI-23-4
c	40	20	4.2	664	13	BM228469	BM228469 K0261G09-
	41	20	4.2	677	10	BB399941	BB399941 BB399941
	42	20	4.2	701	17	BH299972	BH299972 CH230-36D
c	43	20	4.2	909	13	BI739973	BI739973 603361223
	44	20	4.2	1058	12	BF143661	BF143661 601789726
	45	19	4.0	270	14	F19925	F19925 ATTS5964 AC

ALIGNMENTS

RESULT 1
BE370506
LOCUS BE370506 252 bp mRNA linear EST 21-JUL-2000
DEFINITION 601218294F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3587470 5', mRNA sequence.
ACCESSION BE370506
VERSION BE370506.1 GI:9315869
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 252)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8750 row: k column: 23
High quality sequence stop: 242.
FEATURES Location/Qualifiers
source 1. .252

```
                     /organism="Mus musculus"
                     /strain="C2ECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3587470"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       75 a     61 c     43 g     73 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 10;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 3.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        255 ACTTTTGCAAATAAAATATAA 275
              |||||||||||||||||||||
Db        232 ACTTTTGCAAATAAAATATAA 252


RESULT 2
BE147853
LOCUS       BE147853                426 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  RC3-HT0230-201199-013-g05 HT0230 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE147853
VERSION     BE147853.1  GI:8610577
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC3-HT0230-201
            199-013-g05&t3=1999-11-20&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 105
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0230"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      109 a    105 c    121 g     91 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 10;  Length 426;
  Best Local Similarity   100.0%;  Pred. No. 4.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        449 CTCAAAAACAACACACAACCA 469
              |||||||||||||||||||||
Db        406 CTCAAAAACAACACACAACCA 426


RESULT 3
BE205326/c
LOCUS       BE205326                587 bp    mRNA    linear   EST 05-SEP-2000
DEFINITION  EST398002 KV0 Medicago truncatula cDNA clone pKV0-21E10, mRNA
            sequence.
ACCESSION   BE205326
VERSION     BE205326.1  GI:8748622
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
  TITLE     ESTs from uninoculated seedling roots of Medicago truncatula
  JOURNAL   Unpublished (1999)
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            Texas A&M University name:T265098e
            TIGR sequence name:MTGBL29TK
            More information is available at. .
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV0-21E10"
                     /clone_lib="KV0"
                     /tissue_type="Seedling roots"
                     /dev_stage="Immediately prior to inoculation with
                     Sinorhizobium meliloti (0 hour)"
                     /lab_host="E.coli strain XLOLR"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      164 a     69 c    143 g    211 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 10;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        454 AAACAACACACAACCAATCAT 474
              |||||||||||||||||||||
Db         54 AAACAACACACAACCAATCAT 34
```

```
RESULT 4
BI311957/c
LOCUS       BI311957                669 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  EST5313707 GESD Medicago truncatula cDNA clone pGESD16D21 5' end,
            mRNA sequence.
ACCESSION   BI311957
VERSION     BI311957.1  GI:14986284
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho
            ,J. and Fraser,C.M.
  TITLE     ESTs from developing reproductive tissues of Medicago truncatula
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Michael A. Grusak
            USDA/ARS Children's Nutrition Research Center
            Baylor College of Medicine
            1100 Bates Street, Houston, TX 77030-2600, USA
            Tel: 713-798-7044
            Fax: 713-798-7078
            Email: mgrusak@bcm.tmc.edu
            B398812e
             TIGR sequence name: MTPBE23TK
             More information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Medicago truncatula"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pGESD16D21"
                     /clone_lib="GESD"
                     /tissue_type="immature seeds"
                     /dev_stage="Immature seeds, 11 to 19 days after
                     pollination"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Immature seeds, collected from pods ranging in age
                     from 11 to 19 days  after pollination, were harvested from
                     greenhouse-grown plants.  Seeds  were removed and
                     separated from pod walls and were immediately frozen  in
                     liquid nitrogen.  Seeds throughout the age range were
                     pooled for  mRNA extraction. cDNA was prepared from polyA+
                     enriched RNA.  The cDNA was directionally  ligated into
                     the Unizap XR vector from Stratagene and packaged using
                     Gigapack III Gold packaging extracts.  Plasmids containing
                     cDNA inserts were excised from the recombinant lambda-Zap
                     phage using  Ex-assist helper phage and propogated in
                     XLOLR cells."
BASE COUNT      178 a     86 c    182 g    223 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 13;  Length 669;
  Best Local Similarity   100.0%;  Pred. No. 5.2;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        454 AAACAACACACAACCAATCAT 474
              |||||||||||||||||||||
Db        571 AAACAACACACAACCAATCAT 551


RESULT 5
AG143540
LOCUS       AG143540                769 bp    DNA     linear   GSS 08-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-004E06.TJ, genomic survey
            sequence.
ACCESSION   AG143540
VERSION     AG143540.1  GI:16673218
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-004E06.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 769)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .769
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-004E06.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      185 a    203 c    109 g    270 t      2 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 17;  Length 769;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        241 CTACATTAATTACAACTTTTG 261
              |||||||||||||||||||||
Db        103 CTACATTAATTACAACTTTTG 123


RESULT 6
BH371846
LOCUS       BH371846                822 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17
            , DNA sequence.
ACCESSION   BH371846
VERSION     BH371846.1  GI:17317971
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 822)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-162M17.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
```

```
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .822
                      /organism="Anopheles gambiae"
                      /strain="PEST"
                      /db_xref="taxon:7165"
                      /clone="AG-ND-162M17"
                      /clone_lib="ND-TAM"
                      /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      235 a    142 c    155 g    290 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 17;  Length 822;
  Best Local Similarity   100.0%;  Pred. No. 5.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        256 CTTTTGCAAATAAAATATAAA 276
              |||||||||||||||||||||
Db        139 CTTTTGCAAATAAAATATAAA 159


RESULT 7
BF573809
LOCUS       BF573809                845 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  602132006F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4271376 5',
            mRNA sequence.
ACCESSION   BF573809
VERSION     BF573809.1  GI:11647521
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1083  row: l  column: 01
            High quality sequence stop: 165.
FEATURES              Location/Qualifiers
     source           1. .845
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4271376"
                      /clone_lib="NIH_MGC_81"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                      (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                      (ggccattatggcc); 5' and 3' adaptors were used in cloning
                      as follows: 5' adaptor sequence:  5'-CACGGCCATTATGGCC-3'
                      and 3' adaptor sequence:
                      5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                      C, or G and N = A, C,  G, or T).   Average insert size
                      1.55 kb (range 1.0-4.0 kb).  15/15 colonies contained
                      inserts by PCR. This library was enriched for full-length
                      clones and was constructed by Clontech  Laboratories (Palo
                      Alto, CA)."
BASE COUNT      261 a    118 c    241 g    225 t
ORIGIN

  Query Match             4.4%;  Score 21;  DB 12;  Length 845;
  Best Local Similarity   100.0%;  Pred. No. 5.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        337 TTCCACAAAGGCTTAGACTTT 357
              |||||||||||||||||||||
Db        394 TTCCACAAAGGCTTAGACTTT 414


RESULT 8
CNS030N2
LOCUS       CNS030N2                970 bp    DNA     linear   GSS 15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            203E01 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL222455
VERSION     AL222455.1  GI:7881274
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 970)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 970)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 970)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES              Location/Qualifiers
     source           1. .970
                      /organism="Tetraodon nigroviridis"
                      /db_xref="taxon:99883"
                      /clone="203E01"
                      /clone_lib="G"
                      /note="Genoscope sequence ID : C0AG203AC01LP1-end : T7"
BASE COUNT      133 a    334 c    342 g    145 t     16 others
ORIGIN

  Query Match             4.4%;  Score 21;  DB 17;  Length 970;
  Best Local Similarity   100.0%;  Pred. No. 5.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         32 TGCTGCCGAGGTGCTGTCGAA 52
              |||||||||||||||||||||
Db        480 TGCTGCCGAGGTGCTGTCGAA 500
```

```
RESULT 9
BB818014
LOCUS       BB818014                401 bp    mRNA    linear   EST 19-NOV-2001
DEFINITION  BB818014 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus
            musculus cDNA clone G730042D17 3', mRNA sequence.
ACCESSION   BB818014
VERSION     BB818014.1  GI:16990643
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .401
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G730042D17"
                     /clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC
                     cDNA"
                     /tissue_type="lung"
                     /cell_line="RCB-0558 LLC"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                     ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_type=stroma cell,
                     cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
                     cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT      132 a     86 c     57 g    126 t
ORIGIN

  Query Match             4.2%;  Score 20;  DB 10;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        413 AATACCTGCCCATTCCCCTC 432
              ||||||||||||||||||||
Db         14 AATACCTGCCCATTCCCCTC 33


RESULT 10
BI790095/c
LOCUS       BI790095                401 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  ic93b12.x1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus
            musculus cDNA clone IMAGE:5661262 3', mRNA sequence.
ACCESSION   BI790095
VERSION     BI790095.1  GI:15817820
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ic93b12.y1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            MGI:1947588 This sequence now available from the IMAGE consortium,
            for clone orders contact: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1. .401
                     /organism="Mus musculus"
                     /strain="ICR"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5661262"
                     /clone_lib="Melton Normalized Mixed Mouse Pancreas 1
                     N1-MMS1"
                     /sex="Both for embryonic & newborn, male for adult and
                     adult islet"
                     /dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
                     adult, mixed"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five
                     libraries representing EI0.5/12.5 pancreatic bud, E16.5
                     pancreas,  newborn pancreas, adult pancreas, and adult
                     islets of Langerhans were  seperately constructed using
                     SuperScript Plasmid Library kit (Life  Technologies). cDNA
```

was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an Ecot of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

```
BASE COUNT      127 a      56 c      85 g     133 t
ORIGIN

  Query Match             4.2%;  Score 20;  DB 13;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        413 AATACCTGCCCATTCCCCTC 432
              ||||||||||||||||||||
Db        389 AATACCTGCCCATTCCCCTC 370
```

RESULT 11
AA461949

LOCUS	AA461949 405 bp mRNA linear EST 10-JUN-1997
DEFINITION	vg71c09.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:871408 5', mRNA sequence.
ACCESSION	AA461949
VERSION	AA461949.1 GI:2186840
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 405)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:510888 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 353.

FEATURES	Location/Qualifiers
source	1. .405 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:871408" /clone_lib="Soares mouse NbMH" /sex="male" /tissue_type="heart" /dev_stage="4 weeks" /lab_host="DH10B" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAAGTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

```
BASE COUNT      131 a      88 c      59 g     127 t
ORIGIN

  Query Match             4.2%;  Score 20;  DB 9;  Length 405;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        413 AATACCTGCCCATTCCCCTC 432
              ||||||||||||||||||||
Db         36 AATACCTGCCCATTCCCCTC 55
```

RESULT 12
BB810434

LOCUS	BB810434 408 bp mRNA linear EST 19-NOV-2001
DEFINITION	BB810434 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus musculus cDNA clone G730005F13 3', mRNA sequence.
ACCESSION	BB810434
VERSION	BB810434.1 GI:16983063
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 408)
AUTHORS	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii-,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.

FEATURES	Location/Qualifiers
source	1. .408 /organism="Mus musculus" /db_xref="taxon:10090" /clone="G730005F13"

/clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC cDNA"
/tissue_type="lung"
/cell_line="RCB-0558 LLC"
/note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164), (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA), (cell_line=RCB-0559 K-1 . F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1), (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE), (tissue_type=bladder, cell_line=RCB-0544 MBT-2), (tissue_type=bone marrow, cell_type=stroma cell, cell_line= CRL-2028 SR-4987), (tissue_type=colon, cell_line=RCB-0549 Cle-H3), (tissue_type=kidney, cell_line=CCL-142 RAG), (tissue_type=submandibular gland, cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C, cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)"
BASE COUNT 134 a 89 c 57 g 128 t
ORIGIN

Query Match 4.2%; Score 20; DB 10; Length 408;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 413 AATACCTGCCCATTCCCCTC 432
||||||||||||||||||||
Db 22 AATACCTGCCCATTCCCCTC 41

RESULT 13
AZ327901/c
LOCUS AZ327901 413 bp DNA linear GSS 29-SEP-2000
DEFINITION 1M0051K05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0051K05 R, DNA sequence.
ACCESSION AZ327901
VERSION AZ327901.1 GI:10387286
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 413)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0051 row: K column: 05
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 413.
FEATURES Location/Qualifiers
source 1. .413
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0051K05"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 91 a 100 c 75 g 147 t
ORIGIN

Query Match 4.2%; Score 20; DB 17; Length 413;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 AAATAAAATATAAAGAAAGT 282
||||||||||||||||||||
Db 113 AAATAAAATATAAAGAAAGT 94

RESULT 14
BF225508/c
LOCUS BF225508 418 bp mRNA linear EST 29-DEC-2000
DEFINITION uy41f07.x1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:3662149 3', mRNA sequence.
ACCESSION BF225508
VERSION BF225508.1 GI:11133070
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: uy41f07.y1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

MGI:1422917
High quality sequence stop: 406.
FEATURES Location/Qualifiers
source 1. .418
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3662149"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; transgenic model WNT-1, expression driven by

MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 130 a 60 c 88 g 140 t
ORIGIN

Query Match 4.2%; Score 20; DB 12; Length 418;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 413 AATACCTGCCCATTCCCCTC 432
||||||||||||||||||||
Db 387 AATACCTGCCCATTCCCCTC 368

RESULT 15
BG975401
LOCUS BG975401 420 bp mRNA linear EST 12-JUN-2001
DEFINITION 602843052F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4978620 5', mRNA sequence.
ACCESSION BG975401
VERSION BG975401.1 GI:14363038
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 420)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10975 row: h column: 13
High quality sequence stop: 418.
FEATURES Location/Qualifiers
source 1. .420
/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4978620"
/clone_lib="NCI_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."
BASE COUNT 149 a 87 c 58 g 125 t 1 others
ORIGIN

Query Match 4.2%; Score 20; DB 13; Length 420;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 413 AATACCTGCCCATTCCCCTC 432
||||||||||||||||||||
Db 22 AATACCTGCCCATTCCCCTC 41

Search completed: July 20, 2003, 07:09:58
Job time : 1131 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 03:28:35 ; Search time 1535 Seconds
(without alignments)
9005.754 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		451	94.9	473	6	AX008564	AX008564 Sequence
2		53.4	11.2	240020	2	AC121587	AC121587 Mus muscu
3		53.4	11.2	289723	2	AC122935	AC122935 Mus muscu
4		52	10.9	171361	2	AC103286	AC103286 Rattus no
c 5		50.2	10.6	219342	10	AL732443	AL732443 Mouse DNA
c 6		49.6	10.4	16228	6	AX252126	AX252126 Sequence
c 7		49.6	10.4	16228	6	AX348891	AX348891 Sequence
c 8		48.6	10.2	37973	6	AX347098	AX347098 Sequence
c 9		48.4	10.2	15649	6	AX277940	AX277940 Sequence
c 10		48.4	10.2	15649	6	AX323627	AX323627 Sequence
c 11		47.8	10.1	349980	6	AX344550	AX344550 Sequence
c 12		47.4	10.0	8044	6	AX251380	AX251380 Sequence
13		46.4	9.8	549	3	DDAAC1	X16525 Dictyosteli
c 14		46.4	9.8	7522	6	AX345817	AX345817 Sequence
c 15		46.4	9.8	349980	6	AX344566	AX344566 Sequence
c 16		46.2	9.7	171146	9	AL161450	AL161450 Human DNA
c 17		46	9.7	5856	6	AX345471	AX345471 Sequence
c 18		46	9.7	6665	6	AX277841	AX277841 Sequence
c 19		46	9.7	6665	6	AX323516	AX323516 Sequence
c 20		46	9.7	6665	6	AX344985	AX344985 Sequence
21		46	9.7	101297	2	AC098899	AC098899 Rattus no
c 22		46	9.7	349980	6	AX344556	AX344556 Sequence
c 23		45.6	9.6	12592	6	AX251796	AX251796 Sequence
c 24		45.6	9.6	349980	6	AX344573	AX344573 Sequence
c 25		45.4	9.6	18997	6	AX345473	AX345473 Sequence
c 26		45.4	9.6	18997	6	AX347385	AX347385 Sequence
c 27		45.4	9.6	18997	6	AX349106	AX349106 Sequence
28		45.2	9.5	105783	9	AL162576	AL162576 Human DNA
c 29		45	9.5	6172	6	AX344634	AX344634 Sequence
c 30		45	9.5	8883	6	AX251519	AX251519 Sequence
31		45	9.5	48068	9	AL589983	AL589983 Human DNA
32		45	9.5	216734	5	AB073376	AB073376 Oryzias l
c 33		44.8	9.4	5268	6	AX345779	AX345779 Sequence
c 34		44.8	9.4	17738	6	AX346440	AX346440 Sequence
c 35		44.8	9.4	321003	2	PFMAL4P3	AL035476 Plasmodiu
c 36		44.6	9.4	7352	6	AX345272	AX345272 Sequence
c 37		44.6	9.4	139990	2	AC095599	AC095599 Rattus no
38		44.6	9.4	272410	2	AC107492	AC107492 Rattus no
c 39		44.6	9.4	349980	6	AX344565	AX344565 Sequence
c 40		44.4	9.3	1141	6	AX083744	AX083744 Sequence
c 41		44.4	9.3	349980	6	AX344557	AX344557 Sequence
c 42		44.2	9.3	5768	6	AX347064	AX347064 Sequence
43		44	9.3	49981	6	AX059550	AX059550 Sequence
44		44	9.3	63292	8	F14G16	AF147260 Arabidops
45		44	9.3	152619	2	AC101920	AC101920 Mus muscu

ALIGNMENTS

RESULT 1
AX008564
LOCUS AX008564 473 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO9966057.
ACCESSION AX008564
VERSION AX008564.1 GI:9996114
KEYWORDS .
SOURCE garden asparagus.
ORGANISM Asparagus officinalis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;
Asparagaceae; Asparagus.
REFERENCE 1 (bases 1 to 473)
AUTHORS Draper,J., Kenton,P. and Paul,W.
TITLE Inducible promoters

```
  JOURNAL   Patent: WO 9966057-A 1 23-DEC-1999;
            DRAPER JOHN (GB); KENTON PAUL (GB); BIOGEMMA UK LTD (GB); PAUL
            WYATT (GB)
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Asparagus officinalis"
                     /db_xref="taxon:4686"
BASE COUNT      162 a    112 c     67 g    132 t
ORIGIN

  Query Match             94.9%;  Score 451;  DB 6;  Length 473;
  Best Local Similarity   99.6%;  Pred. No. 4e-93;
  Matches  473;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60

Qy         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120

Qy        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180

Qy        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240

Qy        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300

Qy        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360

Qy        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTG 420
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCC-AATTATCCCTATAAATACCTG 419

Qy        421 CCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATCATG 475
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        420 CCCATTCCCCTCCTCCAGACTCATCTAACT-AAAAACAACACACAACCAATCATG 473


RESULT 2
AC121587
LOCUS       AC121587               240020 bp    DNA     linear   HTG 25-MAY-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-273O2, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC121587
VERSION     AC121587.2  GI:21206457
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 240020)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 240020)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 240020)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On May 25, 2002 this sequence version replaced gi:20986665.

            ---------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0273O02
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 234238 bases at least Q40
            Consensus quality: 235156 bases at least Q30
            Consensus quality: 235749 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 238837; sum-of-contigs
            Quality coverage: 15.75 in Q20 bases; agarose-fp
            Quality coverage: 9.77 in Q20 bases; sum-of-contigs
            ------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1565: contig of 1565 bp in length
            *     1566     1665: gap of unknown length
            *     1666     8142: contig of 6477 bp in length
            *     8143     8242: gap of unknown length
            *     8243    19060: contig of 10818 bp in length
            *    19061    19160: gap of unknown length
            *    19161    40614: contig of 21454 bp in length
            *    40615    40714: gap of unknown length
            *    40715    74844: contig of 34130 bp in length
            *    74845    74944: gap of unknown length
            *    74945    95947: contig of 21003 bp in length
            *    95948    96047: gap of unknown length
            *    96048   127209: contig of 31162 bp in length
            *   127210   127309: gap of unknown length
            *   127310   155476: contig of 28167 bp in length
            *   155477   155576: gap of unknown length
            *   155577   238360: contig of 82784 bp in length
            *   238361   238460: gap of unknown length
            *   238461   239069: contig of 609 bp in length
            *   239070   239169: gap of unknown length
            *   239170   240020: contig of 851 bp in length.
FEATURES             Location/Qualifiers
     source          1. .240020
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-273O2"
     misc_feature    1. .1565
                     /note="assembly_name:Contig56"
     misc_feature    1666. .8142
                     /note="assembly_name:Contig57"
     misc_feature    8243. .19060
                     /note="assembly_name:Contig58"
     misc_feature    19161. .40614
                     /note="assembly_name:Contig59"
     misc_feature    40715. .74844
                     /note="assembly_name:Contig60"
     misc_feature    74945. .95947
                     /note="assembly_name:Contig61"
     misc_feature    96048. .127209
                     /note="assembly_name:Contig62"
```

```
     misc_feature    127310. .155476
                     /note="assembly_name:Contig63"
     misc_feature    155577. .238360
                     /note="assembly_name:Contig64"
     misc_feature    238461. .239069
                     /note="assembly_name:Contig33"
     misc_feature    239170. .240020
                     /note="assembly_name:Contig25"
BASE COUNT    71540 a  49697 c  48772 g  68844 t   1167 others
ORIGIN

  Query Match             11.2%;  Score 53.4;  DB 2;  Length 240020;
  Best Local Similarity   54.3%;  Pred. No. 0.013;
  Matches  108;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        167 AATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAAT 226
              || |  |   | |||| |  |  ||  ||||  ||| |  | |   | || ||||| |  |
Db      67358 AAAAAGAGACAAATGAGATAATCATTGAAATTATTTTCCTAATATTGTGTGTGTACATTT 67417

Qy        227 TGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTA 286
              || | |  | |||  ||  ||||    |  |  |||| || |  | |||  | | ||  |
Db      67418 TGTACAGAGGATATTTAATTTAAAATTATTTCATGCACATTATTTTTAAGAATAATTGAA 67477

Qy        287 ACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAG 346
              | |  || | || |  ||| |  || | ||    |  |||    |  | |  ||| | ||
Db      67478 ATAATAATAATAATAATAATAATAATAATACAAGATCAACGTACAAGTAAACCCAGACAG 67537

Qy        347 GCTTAGACTTTCCACAAAT 365
              || ||| |  | || ||||
Db      67538 GCGTAGTCACTGCAGAAAT 67556


RESULT 3
AC122935
LOCUS       AC122935              289723 bp    DNA     linear   HTG 29-MAY-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-20E4, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC122935
VERSION     AC122935.1  GI:21218568
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 289723)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 289723)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 289723)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0020E04
            ------------------------------------------------------
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 285694 bases at least Q40
            Consensus quality: 286828 bases at least Q30
            Consensus quality: 287381 bases at least Q20
            Insert size: 184000; agarose-fp
            Insert size: 294517; sum-of-contigs
            Quality coverage: 21.66 in Q20 bases; agarose-fp
            Quality coverage: 9.89 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3055: contig of 3055 bp in length
            *     3056     3155: gap of unknown length
            *     3156     7772: contig of 4617 bp in length
            *     7773     7872: gap of unknown length
            *     7873    11259: contig of 3387 bp in length
            *    11260    11359: gap of unknown length
            *    11360    16743: contig of 5384 bp in length
            *    16744    16843: gap of unknown length
            *    16844    25333: contig of 8490 bp in length
            *    25334    25433: gap of unknown length
            *    25434    37441: contig of 12008 bp in length
            *    37442    37541: gap of unknown length
            *    37542    73397: contig of 35856 bp in length
            *    73398    73497: gap of unknown length
            *    73498   118271: contig of 44774 bp in length
            *   118272   118371: gap of unknown length
            *   118372   217099: contig of 98728 bp in length
            *   217100   217199: gap of unknown length
            *   217200   289723: contig of 72524 bp in length.
FEATURES             Location/Qualifiers
     source          1. .289723
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-20E4"
     misc_feature    1. .3055
                     /note="assembly_name:Contig19"
     misc_feature    3156. .7772
                     /note="assembly_name:Contig20"
     misc_feature    7873. .11259
                     /note="assembly_name:Contig21"
     misc_feature    11360. .16743
                     /note="assembly_name:Contig22"
     misc_feature    16844. .25333
                     /note="assembly_name:Contig23"
     misc_feature    25434. .37441
                     /note="assembly_name:Contig24"
     misc_feature    37542. .73397
                     /note="assembly_name:Contig25"
     misc_feature    73498. .118271
                     /note="assembly_name:Contig26"
     misc_feature    118372. .217099
                     /note="assembly_name:Contig27"
     misc_feature    217200. .289723
                     /note="assembly_name:Contig28"
BASE COUNT    88915 a  57516 c  56681 g  85704 t    907 others
ORIGIN

  Query Match             11.2%;  Score 53.4;  DB 2;  Length 289723;
  Best Local Similarity   54.3%;  Pred. No. 0.013;
  Matches  108;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        167 AATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAAT 226
              || |  |   | |||| | |  ||  ||||  ||| |  | |   | || ||||| |  |
Db      65911 AAAAAGAGACAAATGAGATAATCATTGAAATTATTTTCCTAATATTGTGTGTGTACATTT 65970

Qy        227 TGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTA 286
```

```
Db      65971 TGTACAGAGGATATTTAATTTAAAATTATTTCATGCACATTATTTTTAAGAATAATTGAA 66030

Qy        287 ACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAG 346

Db      66031 ATAATAATAATAATAATAATAATAATAATACAAGATCAACGTACAAGTAAACCCAGACAG 66090

Qy        347 GCTTAGACTTTCCACAAAT 365

Db      66091 GCGTAGTCACTGCAGAAAT 66109


RESULT 4
AC103286
LOCUS       AC103286               171361 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-123H2, *** SEQUENCING IN PROGRESS
            ***, 59 unordered pieces.
ACCESSION   AC103286
VERSION     AC103286.3  GI:21731284
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 171361)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171361)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 171361)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 11, 2002 this sequence version replaced gi:17974798.
            ------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
               Center project name: GJSK
               Center clone name: CH230-123H2
            ------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 141160 bases at least Q40
               Consensus quality: 146679 bases at least Q30
               Consensus quality: 150766 bases at least Q20
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 59 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1048: contig of 1048 bp in length
            *     1049     1148: gap of unknown length
            *     1149     2646: contig of 1498 bp in length
            *     2647     2746: gap of unknown length
            *     2747     3831: contig of 1085 bp in length
            *     3832     3931: gap of unknown length
            *     3932     5306: contig of 1375 bp in length
            *     5307     5406: gap of unknown length
            *     5407     7652: contig of 2246 bp in length
            *     7653     7752: gap of unknown length
            *     7753     8937: contig of 1185 bp in length
            *     8938     9037: gap of unknown length
            *     9038    10439: contig of 1402 bp in length
            *    10440    10539: gap of unknown length
            *    10540    12020: contig of 1481 bp in length
            *    12021    12120: gap of unknown length
            *    12121    13564: contig of 1444 bp in length
            *    13565    13664: gap of unknown length
            *    13665    15262: contig of 1598 bp in length
            *    15263    15362: gap of unknown length
            *    15363    16589: contig of 1227 bp in length
            *    16590    16689: gap of unknown length
            *    16690    17906: contig of 1217 bp in length
            *    17907    18006: gap of unknown length
            *    18007    19745: contig of 1739 bp in length
            *    19746    19845: gap of unknown length
            *    19846    21908: contig of 2063 bp in length
            *    21909    22008: gap of unknown length
            *    22009    23163: contig of 1155 bp in length
            *    23164    23263: gap of unknown length
            *    23264    24452: contig of 1189 bp in length
            *    24453    24552: gap of unknown length
            *    24553    25706: contig of 1154 bp in length
            *    25707    25806: gap of unknown length
            *    25807    27922: contig of 2116 bp in length
            *    27923    28022: gap of unknown length
            *    28023    29653: contig of 1631 bp in length
            *    29654    29753: gap of unknown length
            *    29754    31169: contig of 1416 bp in length
```

```
*    31170    31269: gap of unknown length
*    31270    33415: contig of 2146 bp in length
*    33416    33515: gap of unknown length
*    33516    35710: contig of 2195 bp in length
*    35711    35810: gap of unknown length
*    35811    37964: contig of 2154 bp in length
*    37965    38064: gap of unknown length
*    38065    40086: contig of 2022 bp in length
*    40087    40186: gap of unknown length
*    40187    41776: contig of 1590 bp in length
*    41777    41876: gap of unknown length
*    41877    44079: contig of 2203 bp in length
*    44080    44179: gap of unknown length
*    44180    47089: contig of 2910 bp in length
*    47090    47189: gap of unknown length
*    47190    49139: contig of 1950 bp in length
*    49140    49239: gap of unknown length
*    49240    51249: contig of 2010 bp in length
*    51250    51349: gap of unknown length
*    51350    54113: contig of 2764 bp in length
*    54114    54213: gap of unknown length
*    54214    56805: contig of 2592 bp in length
*    56806    56905: gap of unknown length
*    56906    59181: contig of 2276 bp in length
*    59182    59281: gap of unknown length
*    59282    61685: contig of 2404 bp in length
*    61686    61785: gap of unknown length
*    61786    65961: contig of 4176 bp in length
*    65962    66061: gap of unknown length
*    66062    69179: contig of 3118 bp in length
*    69180    69279: gap of unknown length
*    69280    71949: contig of 2670 bp in length
*    71950    72049: gap of unknown length
*    72050    74860: contig of 2811 bp in length
*    74861    74960: gap of unknown length
*    74961    77963: contig of 3003 bp in length
*    77964    78063: gap of unknown length
*    78064    80491: contig of 2428 bp in length
*    80492    80591: gap of unknown length
*    80592    84117: contig of 3526 bp in length
*    84118    84217: gap of unknown length
*    84218    88646: contig of 4429 bp in length
*    88647    88746: gap of unknown length
*    88747    91528: contig of 2782 bp in length
*    91529    91628: gap of unknown length
*    91629    95378: contig of 3750 bp in length
*    95379    95478: gap of unknown length
*    95479    99361: contig of 3883 bp in length
*    99362    99461: gap of unknown length
*    99462   103099: contig of 3638 bp in length
*   103100   103199: gap of unknown length
*   103200   107930: contig of 4731 bp in length
*   107931   108030: gap of unknown length
*   108031   111476: contig of 3446 bp in length
*   111477   111576: gap of unknown length
*   111577   115661: contig of 4085 bp in length
*   115662   115761: gap of unknown length
*   115762   119386: contig of 3625 bp in length
*   119387   119486: gap of unknown length
*   119487   124224: contig of 4738 bp in length
*   124225   124324: gap of unknown length
*   124325   128353: contig of 4029 bp in length
*   128354   128453: gap of unknown length
*   128454   133606: contig of 5153 bp in length
*   133607   133706: gap of unknown length
*   133707   138517: contig of 4811 bp in length
*   138518   138617: gap of unknown length
*   138618   142807: contig of 4190 bp in length

Query Match            10.9%;  Score 52;  DB 2;  Length 171361;
Best Local Similarity  52.3%;  Pred. No. 0.029;
Matches  115;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60
              | |||| | ||| | | ||  |  ||| || ||| |||    || | | ||| |  | ||
Db     133811 GGATTCGTTTTGTGGCATGTGTGCCTTCTTATGCAGCCATCTTGGTTTAGAACTGGCAGT 133870

Qy         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120
              | |  | ||  | || ||  || ||| |  |   |  | |||     | |   |  | ||
Db     133871 TACTAAGAAGTTGCTTGTTGTTTCTTTAACTTTAATATTCAACCTGAACTATTAAATTAT 133930

Qy        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180
              |||| | | | ||   |     |  || || |  ||| |   || | |   ||| ||
Db     133931 GAGATATGAAAATAGTTTGCAAAGGTATCTAAGACCCATGCTACACTAAGTAAATGAATA 133990

Qy        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGT 220
              ||| |   ||    | |||||  || |  || |  | |||
Db     133991 GACTCTGACATCTTACAAATTTCAAATCATCCTGCCATGT 134030


RESULT 5
AL732443/c
LOCUS       AL732443              219342 bp    DNA     linear   ROD 02-JUL-2002
DEFINITION  Mouse DNA sequence from clone RP23-90H6 on chromosome X, complete
            sequence.
ACCESSION   AL732443
VERSION     AL732443.6  GI:21531574
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 219342)
  AUTHORS   Kay,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 21, 2002 this sequence version replaced gi:21314875.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-90H6 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .219342
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-90H6"
                     /clone_lib="RPCI-23"
BASE COUNT    66074 a  38077 c  38575 g  76615 t
ORIGIN
```

```
  Query Match           10.6%;  Score 50.2;  DB 10;  Length 219342;
  Best Local Similarity  54.6%;  Pred. No. 0.073;
  Matches  100;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       139 AAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAA 198
             |||  ||||  || |   |  |  || ||   ||  || |  | |   |||||| ||  |
Db     69492 AAGGCATTACTTTTAGTTTCCCAGTCTAAAATAAATGGAAAACAAAATGCCACAAAACTA 69433

Qy       199 ATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTT 258
             ||||||  ||||  || |   |  |       | |  || |  ||| ||| |       |
Db     69432 TTTCTAAGTAGTAATTTCCATTGCATTGACAATAGAGATTCACTAGATTTAAGTATGATC 69373

Qy       259 TTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGT 318
               |||||||  |||| ||| | |||| ||||   |   |  |||  ||||||||||
Db     69372 CCATATATAAAATTTAAATAAATTACTAATATGATATCCCATCTTAAACTGAAGACTAGA 69313

Qy       319 CCA 321
             | |
Db     69312 CAA 69310


RESULT 6
AX252126/c
LOCUS       AX252126               16228 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 387 from Patent WO0168911.
ACCESSION   AX252126
VERSION     AX252126.1  GI:15985481
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 16228)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the cell cycle
  JOURNAL   Patent: WO 0168911-A 387 20-SEP-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .16228
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     4652 a    198 c   3601 g   7777 t
ORIGIN

  Query Match           10.4%;  Score 49.6;  DB 6;  Length 16228;
  Best Local Similarity  48.3%;  Pred. No. 0.15;
  Matches  139;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy       131 CATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCA 190
             | | ||||   |||  |   | |     |  ||   || | |||  |    | | ||  |
Db      9241 CTTTACTATTTTAACCAAACTTAAAATTTTTAAAAAAAAAAAAAAAAACTAAATATATATA 9182

Qy       191 CAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAAT 250
             | |||||||    ||   | ||  |   || || ||  |||    | |  ||  | | |
Db      9181 TAAAAAAAATATACATACATTTTCTCCACTAAAACTTCAAAATCTTACAATTATTTAATT 9122

Qy       251 TACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGA 310
                 |    ||| ||| || || | | | ||| ||  |  |   ||   || | | |
Db      9121 ATACTATAACACAATTAATATTTATAAATACTTCAAAATTATATTTTAAAACTTATTTTA 9062

Qy       311 AGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGA 370
             |  | |    |||||  |||    | |||    ||| |    | | | |    | | |
Db      9061 AACCAAACGCGATAACTCATACCTACAATCCCAACACTTTAAAAAACCAAAATAAACAAA 9002

Qy       371 TTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACC 418
             | |    ||  |   ||  | ||||||| ||  | |   |||||  ||
Db      9001 TCACGAAATCAAAAAATAAAAACCATCCTAACCAACATAATAAAACCC 8954


RESULT 7
AX348891/c
LOCUS       AX348891               16228 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 349 from Patent WO0202807.
ACCESSION   AX348891
VERSION     AX348891.1  GI:18614926
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with cell signalling
  JOURNAL   Patent: WO 0202807-A 349 10-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .16228
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     4652 a    198 c   3601 g   7777 t
ORIGIN

  Query Match           10.4%;  Score 49.6;  DB 6;  Length 16228;
  Best Local Similarity  48.3%;  Pred. No. 0.15;
  Matches  139;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy       131 CATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCA 190
             | | ||||   |||  |   | |     |  ||   || | |||  |    | | ||  |
Db      9241 CTTTACTATTTTAACCAAACTTAAAATTTTTAAAAAAAAAAAAAAAAACTAAATATATATA 9182

Qy       191 CAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAAT 250
             | |||||||    ||   | ||  |   || || ||  |||    | |  ||  | | |
Db      9181 TAAAAAAAATATACATACATTTTCTCCACTAAAACTTCAAAATCTTACAATTATTTAATT 9122

Qy       251 TACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGA 310
                 |    ||| ||| || || | | | ||| ||  |  |   ||   || | | |
Db      9121 ATACTATAACACAATTAATATTTATAAATACTTCAAAATTATATTTTAAAACTTATTTTA 9062

Qy       311 AGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGA 370
             |  | |    |||||  |||    | |||    ||| |    | | | |    | | |
Db      9061 AACCAAACGCGATAACTCATACCTACAATCCCAACACTTTAAAAAACCAAAATAAACAAA 9002

Qy       371 TTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACC 418
             | |    ||  |   ||  | ||||||| ||  | |   |||||  ||
Db      9001 TCACGAAATCAAAAAATAAAAACCATCCTAACCAACATAATAAAACCC 8954


RESULT 8
AX347098/c
LOCUS       AX347098               37973 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 2169 from Patent WO0200928.
ACCESSION   AX347098
VERSION     AX347098.1  GI:18494986
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 2169 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .37973
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT    10166 a    620 c   8741 g  18446 t
ORIGIN

  Query Match           10.2%;  Score 48.6;  DB 6;  Length 37973;
  Best Local Similarity  48.9%;  Pred. No. 0.22;
  Matches  158;  Conservative    0;  Mismatches  164;  Indels    1;  Gaps    1;

Qy       130 ACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCC 189
```

```
Db      26628 AAATAACTAAAAAAATCATCAAAAACAAAAAAAAAAAAAAAAATTTCAATATAACACACTA 26569

Qy        190 ACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAA 249

Db      26568 AAAACAAAAAAATACAAAATAATTAACTATAACAAAAATCCGAAAAAAATCCTACTATCA 26509

Qy        250 TTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATG 309

Db      26508 CCGTAA-ACATCAAACTAAATAAAAATTATATTACTTAACATATTACAAAAAAAAAAATC 26450

Qy        310 AAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAG 369

Db      26449 TAACCGAATACAATAACTCGTACCTATAATCCCAACACTTTAAAAAACCGAAAAACGTAA 26390

Qy        370 ATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCC 429

Db      26389 ATCACGAAATCAAAAAATCGAAACCATCCTAACCAACATAATAAAACCCCGTCTCTCCTA 26330

Qy        430 CTCCTCCAGACTCATCTAACTCA 452

Db      26329 AAAATACAAAAATTAATAACGCA 26307


RESULT 9
AX277940/c
LOCUS       AX277940               15649 bp    DNA     linear   PAT 01-NOV-2001
DEFINITION  Sequence 103 from Patent WO0177375.
ACCESSION   AX277940
VERSION     AX277940.1  GI:16605012
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with gene regulation
  JOURNAL   Patent: WO 0177375-A 103 18-OCT-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .15649
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     4082 a    261 c   3731 g   7575 t
ORIGIN

  Query Match           10.2%;  Score 48.4;  DB 6;  Length 15649;
  Best Local Similarity   46.7%;  Pred. No. 0.28;
  Matches  154;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        136 CTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAA 195

Db      11798 CTAAACAACCAAACCTACTCTTCCAACTCCCCAAAAATCAATTTTTTCATAAAAAAAACA 11739

Qy        196 AAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAA 255

Db      11738 ACAAAACTTTAAAACACTAAACTATCTATTCTTAAAATTAAATTAAAAAACTAAAAAATA 11679

Qy        256 CTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACT 315

Db      11678 CATTAACTTTCTAATTTATAACACAATTTAACCACATCCTCAACCTCAAACCTCAAAAAT 11619

Qy        316 AGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATC 375

Db      11618 TTTTCTTTCCATTTTTTTTTTTTAAACAAAATTTCGCTCTATTCCCCAAACTAAAATACA 11559

Qy        376 CCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTC 435

Db      11558 ATAACGATATCTCAACTCACTACAACCTCCGCCTCTAAAATTCAAACAATTCTCCTACTT 11499

Qy        436 CAGACTCATCTAACTCAAAAACAACACACA 465

Db      11498 CAACCTCCTAAATAACTAAAACTACAAATA 11469


RESULT 10
AX323627/c
LOCUS       AX323627               15649 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 115 from Patent WO0192565.
ACCESSION   AX323627
VERSION     AX323627.1  GI:18094375
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with dna transcription
  JOURNAL   Patent: WO 0192565-A 115 06-DEC-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .15649
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     4082 a    261 c   3731 g   7575 t
ORIGIN

  Query Match           10.2%;  Score 48.4;  DB 6;  Length 15649;
  Best Local Similarity   46.7%;  Pred. No. 0.28;
  Matches  154;  Conservative    0;  Mismatches  176;  Indels    0;  Gaps    0;

Qy        136 CTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAA 195

Db      11798 CTAAACAACCAAACCTACTCTTCCAACTCCCCAAAAATCAATTTTTTCATAAAAAAAACA 11739

Qy        196 AAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAA 255

Db      11738 ACAAAACTTTAAAACACTAAACTATCTATTCTTAAAATTAAATTAAAAAACTAAAAAATA 11679

Qy        256 CTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACT 315

Db      11678 CATTAACTTTCTAATTTATAACACAATTTAACCACATCCTCAACCTCAAACCTCAAAAAT 11619

Qy        316 AGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATC 375

Db      11618 TTTTCTTTCCATTTTTTTTTTTTAAACAAAATTTCGCTCTATTCCCCAAACTAAAATACA 11559

Qy        376 CCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTC 435

Db      11558 ATAACGATATCTCAACTCACTACAACCTCCGCCTCTAAAATTCAAACAATTCTCCTACTT 11499

Qy        436 CAGACTCATCTAACTCAAAAACAACACACA 465

Db      11498 CAACCTCCTAAATAACTAAAACTACAAATA 11469


RESULT 11
AX344550/c
LOCUS       AX344550               349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1 from Patent WO0200932.
ACCESSION   AX344550
VERSION     AX344550.1  GI:18492436
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of known genetic parameters within the mhc
  JOURNAL   Patent: WO 0200932-A 1 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="synthetic construct"
```

```
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo
                     sapiens)~Original length of seq 1: 3.673778 <223>~split as
                     follows:~seq 01 0.000.001 TO 0.349.980~seq 02 0.300.001
                     649.980~seq 03 600.001 949.980~seq 04 900.001
                     1.249.980~seq 05 1.200.001 1.549.980~seq 06 1.500.001
                     1.849.980~seq 07 1.800.001 2.149.980~seq 08 2.100.001
                     2.449.980~seq 09 2.400.001 2.749.980~seq 10 2.700.001
                     3.049.980~seq 11 3.000.001 3.349.980~seq 12 3.300.001
                     3.649.980~seq 13 3.600.001 3.673.778 <223>~Original length
                     of seq 2: 3.673778 <223>~split as follows.~seq 14
                     0.000.001 TO 0.349.980~seq 15 0.300.001 649.980~seq 16
                     600.001 949.980~seq 17 900.001 1.249.980~seq 18 1.200.001
                     1.549.980~seq 19 1.500.001 1.849.980~seq 20 1.800.001
                     2.149.980~seq 21 2.100.001 2.449.980~seq 22 2.400.001
                     2.749.980~seq 23 2.700.001 3.049.980~seq 24 3.000.001
                     3.349.980~seq 25 3.300.001 3.649.980~seq 26 3.600.001
                     3.673.778"
BASE COUNT   100428 a   4240 c  79243 g 166069 t
ORIGIN

  Query Match           10.1%;  Score 47.8;  DB 6;  Length 349980;
  Best Local Similarity   47.0%;  Pred. No. 0.24;
  Matches  148;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        160 AAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTG 219
              |||| |||| ||| | | | | |   | | ||||||    || |
Db     167784 AAAAAAAAATAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167725

Qy        220 TAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAA 279
              |  ||    |||   ||    || ||||| || || |     ||| |||| | ||| ||
Db     167724 AAACAAAAAAAAAAAAAAATAATAAATTAAATAAAAATAAAAAAAAAAAAAAAAAAAAA 167665

Qy        280 AGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTC 339
              |  | || |  || |  |    || |  ||  | | |  | ||||| |||||
Db     167664 AAATTAAAAAAAAAAAAAAAACGAAAAAAAAAAACCAAATATAATAACTCATACCTACAAT 167605

Qy        340 CACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCATCCA 399
              | |||    ||| |     | | |   | || ||   ||  |   |   | ||  |||
Db     167604 CCCAACACTTTAAAAAACTAAAACAAACAAATCATAAAATCAAAAAACAAAAACTTTCCT 167545

Qy        400 AATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAA 459
              || || |   |||||  || | | | |     | || |   ||  |     |   | |
Db     167544 AACTAACATAATAAAACCCCGTCTCTACTAAAAATACAAAAAAATTAACTAAACGTAATA 167485

Qy        460 CACACAACCAATCAT 474
                |  ||| || ||
Db     167484 ACAAATACCTATAAT 167470


RESULT 12
AX251380/c
LOCUS       AX251380               8044 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 348 from Patent WO0168912.
ACCESSION   AX251380
VERSION     AX251380.1  GI:15984803
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 8044)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with tumor suppressor genes and
            oncogenes
  JOURNAL   Patent: WO 0168912-A 348 20-SEP-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .8044
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     2017 a    211 c   1923 g   3893 t
ORIGIN

  Query Match           10.0%;  Score 47.4;  DB 6;  Length 8044;
  Best Local Similarity   47.9%;  Pred. No. 0.53;
  Matches  168;  Conservative    0;  Mismatches  181;  Indels    2;  Gaps    1;

Qy        124 AGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGAC 183
              | || || |  |  ||| ||  ||    |     || |||     || |||  | || ||
Db       1413 AAAATCATACAATAAAATTATATAAAAAATATTACTCAAAAAATTTAAAAAAAAAATAAC 1354

Qy        184 ACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTA 243
              |  |  | |  ||| ||  | || | || ||   | ||  | |    |  | ||     |
Db       1353 ATTTTAAAAATAAATATATTTATAAATCCTTATATTTAATACTATTCATATCAAACAAAA 1294

Qy        244 CATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCT 303
              || ||    |||  | |  |  |||||||  ||      |   ||     | |  | |  |
Db       1293 CAATATAACCAAACTATCTACCTAAAATAAAACTTCTATAAAAATCATTATAAAATTAAT 1234

Qy        304 AACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAA 363
              |||  |  | | | | |  |||||  |||     |||||    ||| |    | | |
Db       1233 AACTTATAACCGAATACGATAACTCACGCCTATAATCACAACACTTTAAAAAACCGAAAT 1174

Qy        364 ATCGAGATTATCCCATGGACTGA--TGGACACCATCCAAATTATCCCTATAAATACCTGC 421
              |  | |||| |  |   |  |  | | || |||| ||  | |   | |||  |||
Db       1173 AAACAAATTACCTAAAATAAAAAATTTAAAACTATCCTAACCAACATAAAAAAACCCTAT 1114

Qy        422 CCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATC 472
              |  | |      | || | | | | || |  || |  |||||    ||||
Db       1113 CTCTACTAAAAATACAAAATTAACCAAATATAATAACACACACCTATAATC 1063


RESULT 13
DDAAC1
LOCUS       DDAAC1                   549 bp    mRNA    linear   INV 05-AUG-1995
DEFINITION  Dictyostelium discoiedeum AAC-rich mRNA (AAC1).
ACCESSION   X16525
VERSION     X16525.1  GI:7172
KEYWORDS    threonine rich protein.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Shaw,D.R., Richter,H., Giorda,R., Ohmachi,T. and Ennis,H.L.
  TITLE     Nucleotide sequences of Dictyostelium discoideum developmentally
            regulated cDNAs rich in (AAC) imply proteins that contain clusters
            of asparagine, glutamine, or threonine
  JOURNAL   Mol. Gen. Genet. 218 (3), 453-459 (1989)
  MEDLINE   90066348
   PUBMED   2511421
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Dictyostelium discoideum"
                     /db_xref="taxon:44689"
                     /clone="AAC1"
                     /clone_lib="lambda gt10"
     CDS             <1. .>549
                     /note="coding region (AA 1 - 183)"
                     /codon_start=1
                     /protein_id="CAA34532.1"
                     /db_xref="GI:930011"
                     /db_xref="SWISS-PROT:P14195"
                     /translation="DNDTNSSTRPNKQQKLLKSNESTTSTTTTTTPITTTTTTTTTTT
                     TPNLSKYYNYNLYIEKQNEKQNLPTTETETTTIITPTLTTTTTTTTTKQIQNTTTSTIN
                     HFLPLIIQKEIIFLLVELGSFLNARKVCKYWKKVCNGCVENLNIYFTDIHLSASVKHV
                     SEVFVKSLNSDYFHLQSVSFING"
     misc_feature    67. .135
                     /note="threonine cluster"
     misc_feature    204. .270
                     /note="threonine cluster"
BASE COUNT      264 a     98 c     45 g    142 t
ORIGIN

  Query Match           9.8%;  Score 46.4;  DB 3;  Length 549;
  Best Local Similarity   47.9%;  Pred. No. 1.4;
```

```
  Matches  134;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        100 CAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCT 159
              ||| || ||||   ||  ||||  ||  || | |  ||    |  | |   | |
Db         40 CAAAAACTATTAAAATCAAATGAATCAACAACTTCAACAACAACAACAACCACACCGATA 99

Qy        160 AAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTG 219
              | |||   |  | | |      ||| | |||  |  || ||||   |    ||
Db        100 ACAACAACAACAACAACAACAACAACAACAACACCAAATCTATCAAAATATTATAAT 159

Qy        220 TAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAA 279
              || || ||  | | |||| | | | || ||  ||||  ||| | || || | | | ||||
Db        160 TATAATTTATATATTGAAAAACAAAATGAAAAACAAAATTTACCAACAACAGAAACAGAA 219

Qy        280 AGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTC 339
              | | | | ||| |   |||     |  |  |  || |   ||  |||    |    | |
Db        220 ACTACCACCATCACACCTACACTAACAACAACAACAACAACAACAACAAAACAAATA 279

Qy        340 CACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCAT 379
              || ||      |   |   ||  ||   |  || | ||||
Db        280 CAAAATACAACAACATCAACAATAAATCATTTTCTACCAT 319


RESULT 14
AX345817/c
LOCUS       AX345817               7522 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 888 from Patent WO0200928.
ACCESSION   AX345817
VERSION     AX345817.1  GI:18493703
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 888 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .7522
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     2074 a     79 c   1578 g   3791 t
ORIGIN

  Query Match             9.8%;  Score 46.4;  DB 6;  Length 7522;
  Best Local Similarity   53.3%;  Pred. No. 0.91;
  Matches   98;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        103 TAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAA 162
              |||   | ||| | |||| | |   |   | | |  ||   ||  | |  |||| |||||
Db       3896 TAACATTCTCCTTTTCATAACACTATATCTTCAAAAACAAAAAAATCGAAATCTCCTAAT 3837

Qy        163 ACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAG 222
                |||| |  | | | || | ||||| | | | |   ||  ||  |    ||   |
Db       3836 TTTCAAAAATATCTATATTAAACATCAAAATATATTCTTTAAAACAAATCTTTAATAATC 3777

Qy        223 AAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGT 282
              | ||   |||   || || | ||| |  |   |   |   ||| ||||||| || |  ||
Db       3776 ATATAACAAAAAAAAAACTTTCATAATCTTATAACATAAAAAAAAAAATATTAAAACCGT 3717

Qy        283 TCTA 286
              ||||
Db       3716 TCTA 3713


RESULT 15
AX344566/c
LOCUS       AX344566              349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 17 from Patent WO0200932.
ACCESSION   AX344566
VERSION     AX344566.1  GI:18492452
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of known genetic parameters within the mhc
  JOURNAL   Patent: WO 0200932-A 17 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo
                     sapiens)~Original length of seq 1: 3.673778 <223>~split as
                     follows.~seq 01 0.000.001 TO 0.349.980~seq 02 0.300.001
                     649.980~seq 03 600.001 949.980~seq 04 900.001
                     1.249.980~seq 05 1.200.001 1.549.980~seq 06 1.500.001
                     1.849.980~seq 07 1.800.001 2.149.980~seq 08 2.100.001
                     2.449.980~seq 09 2.400.001 2.749.980~seq 10 2.700.001
                     3.049.980~seq 11 3.000.001 3.349.980~seq 12 3.300.001
                     3.649.980~seq 13 3.600.001 3.673.778 <223>~Original length
                     of seq 2: 3.673778 <223>~split as follows.~seq 14
                     0.000.001 TO 0.349.980~seq 15 0.300.001 649.980~seq 16
                     600.001 949.980~seq 17 900.001 1.249.980~seq 18 1.200.001
                     1.549.980~seq 19 1.500.001 1.849.980~seq 20 1.800.001
                     2.149.980~seq 21 2.100.001 2.449.980~seq 22 2.400.001
                     2.749.980~seq 23 2.700.001 3.049.980~seq 24 3.000.001
                     3.349.980~seq 25 3.300.001 3.649.980~seq 26 3.600.001
                     3.673.778"
BASE COUNT   102257 a   3636 c   73499 g   170588 t
ORIGIN

  Query Match             9.8%;  Score 46.4;  DB 6;  Length 349980;
  Best Local Similarity   53.3%;  Pred. No. 0.5;
  Matches   98;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        103 TAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAA 162
              |||   | ||| | |||| | |   |   | | |  ||   ||  | |  |||| |||||
Db      82540 TAACATTCTCCTTTTCATAACACTATATCTTCAAAAACAAAAAAATCGAAATCTCCTAAT 82481

Qy        163 ACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAG 222
                |||| |  | | | || | ||||| | | | |   ||  ||  |    ||   |
Db      82480 TTTCAAAAATATCTATATTAAACATCAAAATATATTCTTTAAAACAAATCTTTAATAATC 82421

Qy        223 AAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGT 282
              | ||   |||   || || | ||| |  |   |   |   ||| ||||||| || |  ||
Db      82420 ATATAACAAAAAAAAAAACTTTCATAATCTTATAACATAAAAAAAAAAATATTAAAACCGT 82361

Qy        283 TCTA 286
              ||||
Db      82360 TCTA 82357


Search completed: July 20, 2003, 05:19:11
Job time : 1539 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 03:02:54 ; Search time 191 Seconds
(without alignments)
5600.519 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	475	100.0	475	21	AAZ29510	A.officinalis thau
c	2	49.6	10.4	16228	24	ABL70459	Chemically treated
c	3	49.6	10.4	16228	24	AAS61424	Human gene regulat
c	4	48.6	10.2	37973	24	ABL34196	Human immune syste
c	5	48.4	10.2	15649	22	AAS45396	Chemically pretrea
c	6	48.4	10.2	15649	24	ABK28241	DNA transcription
c	7	47.4	10.0	8044	22	AAS46626	Tumour suppressor
c	8	46.4	9.8	7522	24	ABL32915	Human immune syste
c	9	46	9.7	5856	24	ABL32569	Human immune syste
c	10	46	9.7	6665	22	AAS45299	Chemically pretrea
c	11	46	9.7	6665	24	ABL32083	Human immune syste
c	12	46	9.7	6665	24	ABK28130	DNA transcription
c	13	45.6	9.6	12592	24	AAS61101	Human gene regulat
c	14	45.4	9.6	18997	24	ABK33949	Human DNA for stag
c	15	45.4	9.6	18997	24	ABL32571	Human immune syste
c	16	45	9.5	6172	24	ABN80042	Human chemically m
c	17	45	9.5	8883	22	AAS46761	Tumour suppressor
c	18	44.8	9.4	5268	24	ABL32877	Human immune syste
c	19	44.8	9.4	17738	24	ABL33538	Human immune syste
c	20	44.6	9.4	7352	24	ABL32370	Human immune syste
c	21	44.2	9.3	5768	24	ABL34162	Human immune syste
	22	44	9.3	163319	21	AAF22306	Arabidopsis thalia
c	23	43.8	9.2	9646	24	ABL33688	Human immune syste
c	24	43.6	9.2	6162	24	ABL92315	Chemically treated
c	25	43.6	9.2	14253	24	ABL33495	Human immune syste
c	26	43.4	9.1	5557	24	ABL33547	Human immune syste
c	27	43.4	9.1	8622	24	ABL34143	Human immune syste
c	28	43.4	9.1	11944	24	ABL34187	Human immune syste
c	29	43.2	9.1	11812	22	AAS45502	Chemically pretrea
c	30	43.2	9.1	11812	22	AAS46742	Tumour suppressor
c	31	43.2	9.1	11812	24	ABL34119	Human immune syste
c	32	43.2	9.1	11812	24	ABK28432	DNA transcription
c	33	42.8	9.0	6694	24	ABL33665	Human immune syste
c	34	42.8	9.0	513445	22	AAI61373	Soybean 318013 reg
c	35	42.6	9.0	3920	24	ABQ67030	Human angiogenesis
c	36	42.6	9.0	6290	24	ABL33047	Human immune syste
c	37	42.6	9.0	10717	24	ABN80210	Human chemically m
c	38	42.6	9.0	10717	24	ABL33694	Human immune syste
c	39	42.6	9.0	33053	24	ABQ67006	Human angiogenesis
c	40	42.4	8.9	5958	24	ABL33059	Human immune syste
c	41	42.4	8.9	6866	24	ABL49320	Human polynucleoti
c	42	42.4	8.9	6866	24	ABL32667	Human immune syste
c	43	42.4	8.9	15732	22	AAS45389	Chemically pretrea
c	44	42.4	8.9	15732	24	ABK28234	DNA transcription
c	45	42.4	8.9	83391	24	ABQ67094	Human angiogenesis

ALIGNMENTS

RESULT 1
AAZ29510
ID AAZ29510 standard; DNA; 475 BP.
XX
AC AAZ29510;
XX
DT 14-MAR-2000 (first entry)
XX
DE A.officinalis thaumatin-like PR-5 related gene (AoPRT-L) promoter.
XX
KW Inducible promoter; Thaumatin-like PR-5 related gene; AoPRT-L; trait;
KW non-phytoxic inducing agent; Salicylic acid; SA; BTH; transformed plant;
KW systemic activation; developmental; environmental; pathogen resistance;
KW heterologous gene; disease control; sterility; fertility;
KW fruit ripening; ds.
XX
OS Asparagus officinalis.
XX
FH Key Location/Qualifiers
FT misc_feature 100..111
FT /*tag= a
FT /note= "Homologous to Potato wound induced promoter
FT sequence"
FT misc_feature 205..220
FT /*tag= b
FT /note= "Homologous to Tobacco PR-2 promoter sequence"
FT misc_signal 226..341
FT /*tag= c
FT /label= SA_responsive_element
FT /note= "Induced by non-phytoxic agents like Salicylic
FT acid or BTH"
FT misc_feature 251..264

```
FT                   /*tag=  d
FT                   /note= "Homologous to Carrot PR-3 and PR-4 promoter
FT                         sequence"
FT   repeat_region   281..318
FT                   /*tag=  e
FT                   /rpt_type= TANDEM
FT   repeat_unit     281..298
FT                   /*tag=  f
FT                   /note= "18 bp repeat"
FT   TATA_signal     409..416
FT                   /*tag=  g
XX
PN   WO9966057-A2.
XX
PD   23-DEC-1999.
XX
PF   21-JUN-1999;   99WO-GB01949.
XX
PR   19-JUN-1998;   98GB-0013345.
XX
PA   (BIOG-) BIOGEMMA UK LTD.
XX
PI   Draper J,  Kenton P,  Paul W;
XX
DR   WPI; 2000-106107/09.
XX
PT   Novel promoters used to control the expression of heterologous genes in
PT   transformed plants  -
XX
PS   Claim 4; Fig 6; 67pp; English.
XX
CC   The present DNA sequence is a novel inducible promoter, derived from
CC   Asparagus officinalis thaumatin-like PR-5 related gene, AoPRT-L. The
CC   promoter sequence has similarities with other PR promoters. The
CC   promoter is responsive to low levels of an environmentally-acceptable
CC   and non-phytoxic inducing agents, like Salicylic acid or BTH. The
CC   promoters also exhibit low levels of pathogen induced systemic
CC   activation and environmentally or developmentally induced expression.
CC   The inducible promoter sequence is used to control the expression of
CC   heterologous genes in transformed plants, especially genes whose
CC   products affect a trait of the plant, such as pathogen resistance,
CC   disease control, sterility, fertility or fruit ripening.
XX
SQ   Sequence 475 BP; 163 A; 113 C; 67 G; 132 T; 0 other;

  Query Match             100.0%;  Score 475;  DB 21;  Length 475;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-113;
  Matches  475;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATTCTTATTGCGACCTGACTCTCTTGTTGTGCTGCCGAGGTGCTGTCGAAATTTCTGT 60

Qy         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCAT 120

Qy        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGAT 180

Qy        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATAC 240

Qy        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGT 300

Qy        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCA 360

Qy        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTG 420

Qy        421 CCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATCATG 475
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATCATG 475


RESULT 2
ABL70459/c
ID   ABL70459 standard; DNA; 16228 BP.
XX
AC   ABL70459;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated cell signalling DNA sequence#175.
XX
KW   Cell signalling; cytosine methylation; cell signalling disease;
KW   cancer; tumour; cytostatic; ds.
XX
OS   Unidentified.
XX
PN   WO200202807-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP07471.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
XX
PT   Nucleic acid, useful for diagnosis and therapy of diseases associated
PT   with cell signalling e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with cell signalling  -
XX
PS   Claim 1; SEQ ID NO 349; 24pp+sequence listing; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of at least
CC   18 bases of a segment of chemically pretreated DNA of genes associated
CC   with cell signalling. The activity of the modified sequences of the
CC   invention may be described as cytostatic. The object of the invention is
CC   to provide the chemically modified DNA of genes associated with cell
CC   signalling, as well as oligonucleotides and/or PNA-oligomers for
CC   detecting cytosine methylations, as well as a method which is
CC   particularly suitable for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with cell signalling. The
CC   chemically modified DNA provided by the invention is useful for diagnosis
CC   and therapy of diseases such as solid tumours and cancer. The sequences
CC   given in records ABL70111-ABL70626 represent chemically pre-treated
CC   genomic DNA's of genes associated with cell signalling.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;

  Query Match             10.4%;  Score 49.6;  DB 24;  Length 16228;
  Best Local Similarity   48.3%;  Pred. No. 0.0064;
  Matches  139;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        131 CATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCA 190
              | | ||||   |||  |   | |     |  ||   || | |||  |    | | ||  |
Db       9241 CTTTACTATTTTAACCAAACTTAAAATTTTTAAAAAAAAAAAAAAAAACTAAATATATATA 9182

Qy        191 CAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAAT 250
```

```
Db       9181 TAAAAAAAATATACATACATTTTCTCCACTAAAACTTCAAAATCTTACAATTATTTAATT 9122

Qy        251 TACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGA 310

Db       9121 ATACTATAACACAATTAATATTTATAAATACTTCAAAATTATATTTTAAAACTTATTTTA 9062

Qy        311 AGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGA 370

Db       9061 AACCAAACGCGATAACTCATACCTACAATCCCAACACTTTAAAAAACCAAAATAAACAAA 9002

Qy        371 TTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACC 418

Db       9001 TCACGAAATCAAAAAATAAAAACCATCCTAACCAACATAATAAAACCC 8954


RESULT 3
AAS61424/c
ID   AAS61424 standard; DNA; 16228 BP.
XX
AC   AAS61424;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human gene regulation-associated gene oligonucleotide #379.
XX
KW   Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW   cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW   asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW   renal disease; Preeclampsia; cardiac allograft vascular disease;
KW   colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW   immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW   nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200177375-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP03968.
XX
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-017470/02.
XX
PT   New nucleic acid sequences from chemically modified genes associated
PT   with gene regulation, useful for analysing cytosine methylations for
PT   diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT   disease
XX
PS   Disclosure; SEQ ID No 387; 26pp; English.
XX
CC   The invention relates to 224 nucleic acid sequences comprising at least
CC   18 bases of a chemically pretreated gene associated with gene regulation
CC   selected from 43 known genes (or complementary sequences). The
CC   chemical pretreatment converts cytosine bases unmethylated at the
CC   5-position to uracil or another base with hybridisation behaviour
CC   dissimilar to cytosine, to enable analysis of cytosine methylations.
CC   The DNA sequences, oligomers (or sets/arrays) and method are
CC   useful in the diagnosis of diseases (or predisposition to diseases)
CC   associated with gene regulation and in therapy of such diseases, by
CC   enabling analysis of the cytosine methylation patterns of such genes,
CC   kits are provided. They are especially useful in diagnosis
CC   and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC   disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC   asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC   preeclampsia, graft versus-host disease. The present sequence is a
CC   sequence included in the sequence data for this specification and is
CC   associated with the human gene regulation-associated genes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;

  Query Match           10.4%;  Score 49.6;  DB 24;  Length 16228;
  Best Local Similarity  48.3%;  Pred. No. 0.0064;
  Matches  139;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy        131 CATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCA 190

Db       9241 CTTTACTATTTTAACCAAACTTAAAATTTTTAAAAAAAAAAAAAAAAACTAAATATATATA 9182

Qy        191 CAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAAT 250

Db       9181 TAAAAAAAATATACATACATTTTCTCCACTAAAACTTCAAAATCTTACAATTATTTAATT 9122

Qy        251 TACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGA 310

Db       9121 ATACTATAACACAATTAATATTTATAAATACTTCAAAATTATATTTTAAAACTTATTTTA 9062

Qy        311 AGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGA 370

Db       9061 AACCAAACGCGATAACTCATACCTACAATCCCAACACTTTAAAAAACCAAAATAAACAAA 9002

Qy        371 TTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACC 418

Db       9001 TCACGAAATCAAAAAATAAAAACCATCCTAACCAACATAATAAAACCC 8954


RESULT 4
ABL34196/c
ID   ABL34196 standard; DNA; 37973 BP.
XX
AC   ABL34196;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2169.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
```

PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2169; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 37973 BP; 10166 A; 620 C; 8741 G; 18446 T; 0 other;

Query Match 10.2%; Score 48.6; DB 24; Length 37973;
Best Local Similarity 48.9%; Pred. No. 0.015;
Matches 158; Conservative 0; Mismatches 164; Indels 1; Gaps 1;

```
Qy      130 ACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCC 189
            | || ||||||  ||| | |  || |    ||||   || |     | | || |||||
Db    26628 AAATAACTAAAAAAATCATCAAAAACAAAAAAAAAAAAAAAAATTTCAATATAACACACTA 26569

Qy      190 ACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAA 249
            | |  |||||     | |   || |     | |||       |   || ||||||  | |
Db    26568 AAAACAAAAAAATACAAAATAATTAACTATAACAAAAATCCGAAAAAAATCCTACTATCA 26509

Qy      250 TTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATG 309
              ||    |  || ||||  | ||  | | | || |    |  || |     || ||
Db    26508 CCGTAA-ACATCAAACTAAATAAAAATTATATTACTTAACATATTACAAAAAAAAAAATC 26450

Qy      310 AAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAG 369
            | | | ||  ||||||||||     | |||    ||| |   | | |||   |
Db    26449 TAACCGAATACAATAACTCGTACCTATAATCCCAACACTTTAAAAAACCGAAAAACGTAA 26390

Qy      370 ATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCC 429
            || |    ||  |   || || ||||||| ||  | |   |||||  || | |  | |
Db    26389 ATCACGAAATCAAAAAATCGAAACCATCCTAACCAACATAATAAAACCCCGTCTCTCCTA 26330

Qy      430 CTCCTCCAGACTCATCTAACTCA 452
              | || |      |||| ||
Db    26329 AAAATACAAAAATTAATAACGCA 26307
```

RESULT 5
AAS45396/c
ID AAS45396 standard; DNA; 15649 BP.
XX
AC AAS45396;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated genomic DNA associated with cell cycle #51.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02945.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX
PS Claim 1; SEQ ID No 101; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
SQ Sequence 15649 BP; 4082 A; 261 C; 3731 G; 7575 T; 0 other;

Query Match 10.2%; Score 48.4; DB 22; Length 15649;
Best Local Similarity 46.7%; Pred. No. 0.013;
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```
Qy      136 CTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAA 195
            |||||  |   | | || ||  | ||  | | | | |       |  || |   | | |
Db    11798 CTAAACAACCAAACCTACTCTTCCAACTCCCCAAAAATCAATTTTTTCATAAAAAAAACA 11739

Qy      196 AAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAA 255
            | ||  ||    |    |    | | |   |  ||| | |||     | | |||  |  |
Db    11738 ACAAAACTTTAAAACACTAAACTATCTATTCTTAAAATTAAATTAAAAAACTAAAAAATA 11679

Qy      256 CTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACT 315
            | ||  |     || |   || | | ||  | ||      | |  || ||| | || | |
Db    11678 CATTAACTTTCTAATTTATAACACAATTTAACCACATCCTCAACCTCAAACCTCAAAAAT 11619

Qy      316 AGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATC 375
              | |       |  |   || ||  |||||   |    || | | | || | | | ||
Db    11618 TTTTCTTTCCATTTTTTTTTTTTAAACAAAATTTCGCTCTATTCCCCAAACTAAAATACA 11559

Qy      376 CCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTC 435
              |  ||     |  || |  | |||  |   ||| ||||    |   | ||||  ||| ||
Db    11558 ATAACGATATCTCAACTCACTACAACCTCCGCCTCTAAAATTCAAACAATTCTCCTACTT 11499

Qy      436 CAGACTCATCTAACTCAAAAACAACACACA 465
            ||  ||| |  |   | ||||| ||| | |
Db    11498 CAACCTCCTAAATAACTAAAACTACAAATA 11469
```

RESULT 6
ABK28241/c
ID ABK28241 standard; DNA; 15649 BP.
XX
AC ABK28241;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated genomic DNA #58.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;

KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP03973.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer -
XX
PS Claim 1; SEQ ID No 115; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 15649 BP; 4082 A; 261 C; 3731 G; 7575 T; 0 other;

Query Match 10.2%; Score 48.4; DB 24; Length 15649;
Best Local Similarity 46.7%; Pred. No. 0.013;
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 135 CTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAA 195
Db 11798 CTAAACAACCAAACCTACTCTTCCAACTCCCCAAAAATCAATTTTTTCATAAAAAAAACA 11739

Qy 196 AAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAA 255
Db 11738 ACAAAACTTTAAAACACTAAACTATCTATTCTTAAAATTAAATTAAAAAACTAAAAAATA 11679

Qy 256 CTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACT 315
Db 11678 CATTAACTTTCTAATTTATAACACAATTTAACCACATCCTCAACCTCAAACCTCAAAAAT 11619

Qy 316 AGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATC 375
Db 11618 TTTTCTTTCCATTTTTTTTTTTTAAACAAAATTTCGCTCTATTCCCCAAACTAAAATACA 11559

Qy 376 CCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTC 435
Db 11558 ATAACGATATCTCAACTCACTACAACCTCCGCCTCTAAAATTCAAACAATTCTCCTACTT 11499

Qy 436 CAGACTCATCTAACTCAAAAACAACACACA 465
Db 11498 CAACCTCCTAAATAACTAAAACTACAAATA 11469

RESULT 7
AAS46626/c
ID AAS46626 standard; DNA; 8044 BP.
XX
AC AAS46626;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #348.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
PS Claim 1; SEQ ID No 348; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the

```
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8044 BP; 2017 A; 211 C; 1923 G; 3893 T; 0 other;

  Query Match             10.0%;  Score 47.4;  DB 22;  Length 8044;
  Best Local Similarity   47.9%;  Pred. No. 0.02;
  Matches  168;  Conservative    0;  Mismatches  181;  Indels    2;  Gaps    1;

Qy        124 AGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGAC 183
Db       1413 AAAATCATACAATAAAATTATATAAAAAATATTACTCAAAAAATTTAAAAAAAAAATAAC 1354

Qy        184 ACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTA 243
Db       1353 ATTTTAAAAATAAATATATTTATAAATCCTTATATTTAATACTATTCATATCAAACAAAA 1294

Qy        244 CATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCT 303
Db       1293 CAATATAACCAAACTATCTACCTAAAATAAAACTTCTATAAAAATCATTATAAAATTAAT 1234

Qy        304 AACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAA 363
Db       1233 AACTTATAACCGAATACGATAACTCACGCCTATAATCACAACACTTTAAAAAACCGAAAT 1174

Qy        364 ATCGAGATTATCCCATGGACTGA--TGGACACCATCCAAATTATCCCTATAAATACCTGC 421
Db       1173 AAACAAATTACCTAAAATAAAAAATTTAAAACTATCCTAACCAACATAAAAAAACCCTAT 1114

Qy        422 CCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCAATC 472
Db       1113 CTCTACTAAAAATACAAAATTAACCAAATATAATAACACACACCTATAATC 1063


RESULT 8
ABL32915/c
ID   ABL32915 standard; DNA; 7522 BP.
XX
AC   ABL32915;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 888.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 888; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 7522 BP; 2074 A; 79 C; 1578 G; 3791 T; 0 other;

  Query Match              9.8%;  Score 46.4;  DB 24;  Length 7522;
  Best Local Similarity   53.3%;  Pred. No. 0.035;
  Matches   98;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        103 TAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAA 162
Db       3896 TAACATTCTCCTTTTCATAACACTATATCTTCAAAAACAAAAAAATCGAAATCTCCTAAT 3837

Qy        163 ACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAG 222
Db       3836 TTTCAAAAATATCTATATTAAACATCAAAATATATTCTTTAAAACAAATCTTTAATAATC 3777

Qy        223 AAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGT 282
Db       3776 ATATAACAAAAAAAAAACTTTCATAATCTTATAACATAAAAAAAAAAATATTAAAACCGT 3717

Qy        283 TCTA 286
Db       3716 TCTA 3713


RESULT 9
ABL32569/c
ID   ABL32569 standard; DNA; 5856 BP.
XX
AC   ABL32569;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 542.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
```

```
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 542; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5856 BP; 1482 A; 177 C; 1495 G; 2702 T; 0 other;

  Query Match             9.7%;  Score 46;  DB 24;  Length 5856;
  Best Local Similarity   46.8%;  Pred. No. 0.042;
  Matches  145;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy        160 AAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTG 219
              ||||   ||   |||  | |  |  | || |    ||||  | || |    |    |
Db       5543 AAAATAAAAACTAAAAAAAAACCAACCTTCAATTCTAAAATCTATACTCCCTTCAATCAAA 5484

Qy        220 TAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAA 279
              ||  || |  |      | ||  |||  |   |    |  | |||| | ||||  || ||
Db       5483 TAATAAATCCATTTCCGAAACACACAAAACCCATCTTTCCTACAAAAATAATAATAACAA 5424

Qy        280 AGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTC 339
              |     || |  ||   |    | |||   || |       |      |  |   || |
Db       5423 AAAAAAAAAAATAAACATCACCCCAACCATAATATCCCAACTTCTTTTTTTTTTTTTAA 5364

Qy        340 CACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCATCCA 399
               |||||  || |  || |    | || | | | ||    |  | |   | ||| |  | ||
Db       5363 AACAAAATCTCACTCTATTACCCAAACTAAAATACAATAACGCCATCTCGACTCACTACA 5304

Qy        400 AATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAA 459
              || | | |||    |   |   |||||| ||| |  ||  ||| |  |   | |||||
Db       5303 AACTCTACCTCCCGAATTCACGCCATTCTCCTACCTCAACCTCGTAAATAACTAAAACTA 5244

Qy        460 CACACAACCA 469
              || | | |||
Db       5243 CAAATACCCA 5234


RESULT 10
AAS45299/c
ID   AAS45299 standard; DNA; 6665 BP.
XX
AC   AAS45299;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated complementary DNA associated with cell cycle #2.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02945.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analysing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle -
XX
PS   Claim 1; SEQ ID No 4; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers.
XX
SQ   Sequence 6665 BP; 1830 A; 86 C; 1243 G; 3506 T; 0 other;

  Query Match             9.7%;  Score 46;  DB 22;  Length 6665;
  Best Local Similarity   51.5%;  Pred. No. 0.043;
  Matches  106;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy         88 ATTTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTA 147
              || | | |  ||  |    || |||||   |   | | || || |  | ||||  |||||
Db       3168 ATATAATAACTCACACCTATAATTCCAACACTCTAAAAAACCAAAATAATAAATCAATTA 3109

Qy        148 GCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATT 207
                  |     | |||  |||   || | | || | | | || |   |    | ||||
Db       3108 AAACCAAAAATTCAAAACCAAACTAACCAACATAATAAAACCCCCCCATCTCTACTAAAA 3049

Qy        208 AGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATA 267
              |  |      |  | ||| |  | | | |||||||| | |    || ||||    || ||
Db       3048 ATACAAAAATTAAATAAACTTAATAATAAATACCTATAATCTCAACTACTTAAAAAACTA 2989

Qy        268 AAATATAAAGAAAGTTCTAACATGAA 293
              ||||| ||| |    |  ||| | ||
Db       2988 AAATAAAAAAATCACTTAAACCTAAA 2963


RESULT 11
ABL32083/c
ID   ABL32083 standard; DNA; 6665 BP.
XX
AC   ABL32083;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 56.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
```

KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 56; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6665 BP; 1830 A; 86 C; 1243 G; 3506 T; 0 other;

Query Match 9.7%; Score 46; DB 24; Length 6665;
Best Local Similarity 51.5%; Pred. No. 0.043;
Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 88 ATTTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTA 147
Db 3168 ATATAATAACTCACACCTATAATTCCAACACTCTAAAAAACCAAAATAATAAATCAATTA 3109

Qy 148 GCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATT 207
Db 3108 AAACCAAAAATTCAAAACCAAACTAACCAACATAATAAAACCCCCCCATCTCTACTAAAA 3049

Qy 208 AGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATA 267
Db 3048 ATACAAAAATTAAATAAACTTAATAATAAATACCTATAATCTCAACTACTTAAAAAACTA 2989

Qy 268 AAATATAAAGAAAGTTCTAACATGAA 293
Db 2988 AAATAAAAAAATCACTTAAACCTAAA 2963

RESULT 12
ABK28130/c
ID ABK28130 standard; DNA; 6665 BP.
XX
AC ABK28130;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated complementary genomic DNA #2.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP03973.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer -
XX
PS Claim 1; SEQ ID No 4; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 6665 BP; 1830 A; 86 C; 1243 G; 3506 T; 0 other;

Query Match 9.7%; Score 46; DB 24; Length 6665;
Best Local Similarity 51.5%; Pred. No. 0.043;
Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 88 ATTTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTA 147
Db 3168 ATATAATAACTCACACCTATAATTCCAACACTCTAAAAAACCAAAATAATAAATCAATTA 3109

Qy 148 GCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATT 207
Db 3108 AAACCAAAAATTCAAAACCAAACTAACCAACATAATAAAACCCCCCCATCTCTACTAAAA 3049

```
Qy        208 AGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATA 267
              |  |      |  | ||| |  | | | |||||||| | |    || ||||    || ||
Db       3048 ATACAAAAATTAAATAAACTTAATAATAAATACCTATAATCTCAACTACTTAAAAAACTA 2989

Qy        268 AAATATAAAGAAAGTTCTAACATGAA 293
              ||||| ||| |    |  ||| | ||
Db       2988 AAATAAAAAAATCACTTAAACCTAAA 2963
```

RESULT 13
AAS61101/c
ID AAS61101 standard; DNA; 12592 BP.
XX
AC AAS61101;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #56.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Claim 1; SEQ ID No 57; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases; by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12592 BP; 3783 A; 195 C; 2736 G; 5878 T; 0 other;

```
  Query Match             9.6%;  Score 45.6;  DB 24;  Length 12592;
  Best Local Similarity   49.7%;  Pred. No. 0.065;
  Matches  144;  Conservative    0;  Mismatches  144;  Indels    2;  Gaps    1;

Qy        182 ACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACC 241
              ||  ||  | | || ||||          |||    |    |||||    ||  | || |
Db       9190 ACTAATTTAAAAAATAAATAAACCAACAACTTAAAATTACAAAATTCCTTACCAACTAAC 9131

Qy        242 TACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACT--AG 299
              ||   | |      |   |   | ||||||||||| |||  ||||  |   | ||| |
Db       9130 TATTCTTAAATATAAAATATTATTTAAAATATAAACAAACCTCTATAAATTATACTTAAA 9071

Qy        300 TTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCC 359
              |  || || |  |||          ||| |   || |   |||||  || |  || ||
Db       9070 CTTAAATATAATCACTATAAATTTTTCTCTTTTTTTTTAAAACAAAATCTCACTCTATCA 9011

Qy        360 ACAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCT 419
              | || | | ||    |  |    | ||| |  | |||| |   |||   ||   |
Db       9010 CCCAAACTAAAATACAATAACGCGATCTCGACTCACTACAAACTCCACCTCCTAAATTCA 8951

Qy        420 GCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACCA 469
              |||||| ||| |  ||  ||| |  |   | ||||| ||| | | | |
Db       8950 CGCCATTCTCCTACCTCAACCTCCTAAATAACTAAAACTACAAATACCTA 8901
```

RESULT 14
ABK33949/c
ID ABK33949 standard; DNA; 18997 BP.
XX
AC ABK33949;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human DNA for staging of Astrocytomas, complement, #16.
XX
KW Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;
KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
KW matrix assisted laser desorption/ionization mass spectrometry.
XX
OS Homo sapiens.
XX
PN WO200202808-A2.
XX
PD 10-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07538.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-171649/22.
XX
PT Novel chemically modified genomic DNA sequences, useful in the
PT characterisation, classification, differentiation, grading, staging,
PT treatment and/or diagnosis of astrocytomas or predisposition to
PT astrocytomas -
XX
PS Claim 1; SEQ ID No 32; 37pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence (I) of at
CC least 18 bases in length of a segment of chemically pre-treated genomic
CC DNA which has any one of the sequences of (ABK33919-ABK34032) or its
CC complement. Also included are an oligonucleotide or peptide nucleic
CC acid (or set thereof) of at least 9 nucleotides which hybridises to (I),

```
CC   primers for (I), probes for detecting cytosine methylation or single-
CC   nucleotide polymorphisms (SNP) in (I), an array of oligomers
CC   or peptide nucleic acids for analysing diseases associated with the
CC   methylation states of the CpG dinucleotides of (I). The array is useful
CC   for determining genetic and/or epigenetic parameters, classification,
CC   differentiation, grading, staging, treatment and/or diagnosis of
CC   astrocytomas, or the predisposition to astrocytomas by analysing cytosine
CC   methylations, involves obtaining a biological sample containing genomic
CC   DNA, extracting the genomic DNA, converting cytosine bases which are
CC   unmethylated at the 5-position, in the genomic DNA sample, to uracil or
CC   another base which is dissimilar to cytosine in terms of hybridisation
CC   behaviour, by chemical treatment and amplifying chemically pre-treated
CC   genomic DNA fragments using the array and a polymerase, where the
CC   amplificates carry a detectable label. The method further involves
CC   identifying methylation status of one or more cytosine positions, and
CC   analysing methylation status of the cytosine positions by reference to
CC   one or more data sets. The genomic DNA is chemically treated by using a
CC   bisulphite, hydrogen sulphite or disulphite. The amplification
CC   step amplifies DNA which is of particular interest in astrocytoma or
CC   brain tissue, based on the specific genomic methylation status of brain
CC   tissues, as opposed to background DNA. The amplificates carry a
CC   fluorescent label or radionuclide. Optionally, the labels of the
CC   amplificates are detachable molecule fragments having a typical mass
CC   which are detected in a mass spectrometer. The fragments of chemically
CC   pre-treated genomic DNA to be amplified, have a single positive or
CC   negative charge for a better detectability in the mass spectrometer.
CC   Preferably, the amplificates or fragments of the amplificates are
CC   detected by matrix assisted laser desorption/ionization mass spectrometry
CC   (MALDI) or using electron spray mass spectrometry (ESI). The
CC   present sequence is one of the chemically pre-treated reference DNA
CC   samples of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 18997 BP; 6146 A; 299 C; 4160 G; 8392 T; 0 other;

  Query Match             9.6%;  Score 45.4;  DB 24;  Length 18997;
  Best Local Similarity   49.0%;  Pred. No. 0.082;
  Matches  121;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy         99 CCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCC 158
              |  ||   |  |||  |   | | | |     |  | | | ||  || ||   ||     |
Db      11885 CAAAACCTATTTTTAAACCCGTTATAACTAAAATTAAATTAACAAACTAAACTACCTTTC 11826

Qy        159 TAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGT 218
              |      ||||| |||| | || ||  |||   | ||||||  ||||| |         |
Db      11825 TCCTCTCCAATAAAAACTAAATAACTTATCATTAAAAAAAAACCTAATAAAAAAACATAT 11766

Qy        219 GTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGA 278
                | ||       | | || |    |  | || ||| | |   |||||| | |||| |
Db      11765 TAATAATAAAACCAATAAAAAAAATTAAAATTTCCAAATCTATTAAATAACAAATAATAA 11706

Qy        279 AAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATT 338
              || |   || || || || | ||| || |  || |  | ||     | |  |   |||
Db      11705 AACTATAAAAATAAAAACAACTTCAAAAAAAAAAAAAAATTCATAATAATTCTTAATTACC 11646

Qy        339 CCACAAA 345
                || |||
Db      11645 ACAAAAA 11639


RESULT 15
ABL32571/c
ID   ABL32571 standard; DNA; 18997 BP.
XX
AC   ABL32571;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 544.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 544; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 18997 BP; 6146 A; 299 C; 4160 G; 8392 T; 0 other;

  Query Match             9.6%;  Score 45.4;  DB 24;  Length 18997;
  Best Local Similarity   49.0%;  Pred. No. 0.082;
  Matches  121;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy         99 CCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCC 158
              |  ||   |  |||  |   | | | |     |  | | | ||  || ||   ||     |
Db      11885 CAAAACCTATTTTTAAACCCGTTATAACTAAAATTAAATTAACAAACTAAACTACCTTTC 11826

Qy        159 TAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGT 218
              |      ||||| |||| | || ||  |||   | ||||||  ||||| |         |
Db      11825 TCCTCTCCAATAAAAACTAAATAACTTATCATTAAAAAAAAACCTAATAAAAAAACATAT 11766

Qy        219 GTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGA 278
                | ||       | | || |    |  | || ||| | |   |||||| | |||| |
Db      11765 TAATAATAAAACCAATAAAAAAAATTAAAATTTCCAAATCTATTAAATAACAAATAATAA 11706

Qy        279 AAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATT 338
              || |   || || || || | ||| || |  || |  | ||     | |  |   |||
Db      11705 AACTATAAAAATAAAAACAACTTCAAAAAAAAAAAAAAATTCATAATAATTCTTAATTACC 11646

Qy        339 CCACAAA 345
                || |||
Db      11645 ACAAAAA 11639


Search completed: July 20, 2003, 04:53:23
Job time : 193 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 03:30:14 ; Search time 44 Seconds
(without alignments)
3310.715 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	38	8.0	1263	4	US-09-302-681-4	Sequence 4, Appli
	2	36.8	7.7	6243	2	US-09-056-075-1	Sequence 1, Appli
c	3	36.4	7.7	4093	4	US-07-861-458C-5	Sequence 5, Appli
c	4	35	7.4	3815	4	US-08-936-165A-196	Sequence 196, App
	5	34.4	7.2	457	4	US-09-643-597-306	Sequence 306, App
c	6	33.8	7.1	24595	6	5428147-1	Patent No. 5428147
c	7	33.6	7.1	4673	1	US-07-638-431-1	Sequence 1, Appli
c	8	33.6	7.1	4673	5	PCT-US92-00018-1	Sequence 1, Appli
c	9	33.6	7.1	9412	2	US-08-955-138-1	Sequence 1, Appli
c	10	33.6	7.1	168575	4	US-09-426-290-1	Sequence 1, Appli
c	11	33.2	7.0	1253	2	US-08-591-629-1	Sequence 1, Appli
c	12	33.2	7.0	8700	2	US-08-392-625-16	Sequence 16, Appl
c	13	33.2	7.0	8700	2	US-08-466-961A-16	Sequence 16, Appl
c	14	33.2	7.0	8700	2	US-08-645-193B-18	Sequence 18, Appl
	15	33	6.9	357	2	US-08-666-405-12	Sequence 12, Appl
	16	33	6.9	575	2	US-08-569-166-29	Sequence 29, Appl
	17	33	6.9	882	2	US-08-666-405-27	Sequence 27, Appl
c	18	32.8	6.9	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	19	32.8	6.9	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	20	32.8	6.9	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	21	32.6	6.9	1001	4	US-09-641-638-598	Sequence 598, App
c	22	32.6	6.9	1001	4	US-09-641-638-599	Sequence 599, App
c	23	32.6	6.9	12793	4	US-09-004-838-124	Sequence 124, App
c	24	32.6	6.9	15062	4	US-09-004-838-89	Sequence 89, Appl
	25	32.6	6.9	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	26	32.4	6.8	359	2	US-08-687-080-65	Sequence 65, Appl
	27	32.4	6.8	5609	4	US-09-313-677-14	Sequence 14, Appl
	28	32.4	6.8	5609	4	US-09-313-677-15	Sequence 15, Appl
	29	32.4	6.8	7026	4	US-09-313-677-20	Sequence 20, Appl
	30	32.4	6.8	7344	4	US-09-313-677-16	Sequence 16, Appl
c	31	32.4	6.8	98844	4	US-09-791-211-10	Sequence 10, Appl
c	32	32	6.7	600	4	US-09-385-982-482	Sequence 482, App
c	33	32	6.7	773	4	US-09-385-982-541	Sequence 541, App
c	34	32	6.7	2639	4	US-09-374-135-1	Sequence 1, Appli
	35	31.8	6.7	246240	2	US-08-724-394A-20	Sequence 20, Appl
	36	31.8	6.7	246240	2	US-08-724-394A-21	Sequence 21, Appl
	37	31.8	6.7	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	38	31.6	6.7	658	4	US-08-998-416-595	Sequence 595, App
	39	31.6	6.7	731	1	US-08-451-405A-2	Sequence 2, Appli
	40	31.6	6.7	3558	4	US-09-134-001C-2439	Sequence 2439, Ap
	41	31.6	6.7	13737	4	US-09-538-414-10	Sequence 10, Appl
c	42	31.4	6.6	568	4	US-08-642-274D-23	Sequence 23, Appl
c	43	31.4	6.6	568	4	US-08-952-014C-23	Sequence 23, Appl
	44	31.4	6.6	2539	3	US-08-749-522-3	Sequence 3, Appli
c	45	31.4	6.6	5099	4	US-09-610-040-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-302-681-4
; Sequence 4, Application US/09302681
; Patent No. 6441149
; GENERAL INFORMATION:
; APPLICANT: Herrnstadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Clevenger, William
; APPLICANT: Fahy, Eoin F.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON
; TITLE OF INVENTION: QUANTIFICATION OF EXTRAMITOCHONDRIAL DNA
; FILE REFERENCE: 660088.416C1
; CURRENT APPLICATION NUMBER: US/09/302,681
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-302-681-4

Query Match 8.0%; Score 38; DB 4; Length 1263;
Best Local Similarity 57.6%; Pred. No. 0.12;
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy        102 ATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAA 161
              || ||||| | ||  | ||| |  ||  | || | |||| | |  || | | ||||  |
Db        803 ATGATTATGTACACATTATGTGCCAATGAAATAAATAAAATTAAAAGGTAATTCCCAAAT 862

Qy        162 AACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTG 219
              || | || ||| | |  | | | | |  | ||| | |  |  ||  | ||||| |||
Db        863 AAATAAAAACAGAAAATTTTAAAAAGAGAGAGAGAGAGATGAGATGGGGCTTTGGGTG 920
```

RESULT 2
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street

```
;       CITY:  Madison
;       STATE:  WI
;       COUNTRY:  US
;       ZIP:  53701-2113
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/056,075
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seay, Nicholas J.
;       REGISTRATION NUMBER:  27386
;       REFERENCE/DOCKET NUMBER:  960296.95238
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  608-251-5000
;       TELEFAX:  608-251-9166
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6243 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3770..4013
;       OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
;       OTHER INFORMATION:  plasmid RP4"
US-09-056-075-1

  Query Match             7.7%;  Score 36.8;  DB 2;  Length 6243;
  Best Local Similarity   49.5%;  Pred. No. 0.45;
  Matches   95;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy        122 AGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATG 181
              | | ||  | |  |||||  || ||    |||     || |   ||   |||  | |
Db       1213 AAAAAAATAAAAAAATAAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAA 1272

Qy        182 ACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACC 241
              | |     | |  |||||    ||  | | |     |  | ||||   ||| | || |
Db       1273 AAAATAAAAAAATAAAAAAATAAAAAAATATAAAAATAAAAAAATATAAAAATAAAAAAA 1332

Qy        242 TACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTT 301
              || |  ||| | ||  | |  |||||||| |  || ||| |    || || || |
Db       1333 TATAAAAATAAAAAAATATAAAAATAAAAAAATAAAAAAATATAAAAATAAAAAAATAAA 1392

Qy        302 CTAACATGAAGA 313
              || || || |
Db       1393 AAAATATAAAAA 1404


RESULT 3
US-07-861-458C-5/c
; Sequence 5, Application US/07861458C
; Patent No. 6232061
;  GENERAL INFORMATION:
;    APPLICANT:  Marchionni, Mark Andrew
;    APPLICANT:  Johnson, Carl D.
;    TITLE OF INVENTION:  HOMOLOGY CLONING
;    NUMBER OF SEQUENCES:  142
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/861,458C
;      FILING DATE:  04/01/92
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  04585/014001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4093
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-861-458C-5

  Query Match             7.7%;  Score 36.4;  DB 4;  Length 4093;
  Best Local Similarity   53.5%;  Pred. No. 0.5;
  Matches   76;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy         33 GCTGCCGAGGTGCTGTCGAAATTTCTGTTGCGCACAACATACTGGTCCTTGCTTGATTTG 92
              |||  ||  ||  |  ||  || | ||   | ||| ||| | | ||  || |||  |  |
Db       1432 GCTTTCGGCGTTGTTCCGCTATCTTTGAGCCACACGACAGAATCGTAGTTTCTTAAAATA 1373

Qy         93 ACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTA 152
              | ||||  |  |||  | |  |  | ||    |||  |   | || ||   | ||| | |
Db       1372 ATAGTTGAATAAATGGTCTTAAAATAATAGTTGAATAAATGGTCTTAAAATAATAGTTGA 1313

Qy        153 ATCCCCTAAAACTCAATACAAA 174
              ||    |||||   | || |||
Db       1312 ATAAAATAAAAGATATTAGAAA 1291


RESULT 4
US-08-936-165A-196/c
; Sequence 196, Application US/08936165A
; Patent No. 6348582
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael
;    APPLICANT:  Burnham, Martin
;    APPLICANT:  Hodgson, John
;    APPLICANT:  Knowles, David
;    APPLICANT:  Lonetto, Michael
;    APPLICANT:  Nicholas, Richard
;    APPLICANT:  Pratt, Julie
;    APPLICANT:  Reichard, Richard
;    APPLICANT:  Rosenberg, Martin
;    APPLICANT:  Ward, Judith
;    TITLE OF INVENTION:  No. 6348582el Prokaryotic Polynucleotides,
;    TITLE OF INVENTION:  Polypeptides and Their Uses
;    NUMBER OF SEQUENCES:  534
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,165A
;      FILING DATE:  24-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,032
;      FILING DATE:  24-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P50549
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  196:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3815 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-936-165A-196

  Query Match             7.4%;  Score 35;  DB 4;  Length 3815;
  Best Local Similarity   59.6%;  Pred. No. 1.2;
  Matches   59;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        194 AAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTAC 253
              | || ||||| |  ||  | ||| | ||  ||||  |||    |  | |||    | | |
Db       2467 ATAATATTCTTAGAAGCATATGCATCTATTAATTTTAAAGAATAGGCGTACGCATAATTC 2408

Qy        254 AACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGA 292
              | |||| ||||||||||| | |  || ||    ||| |
Db       2407 CAATTTTTCAAATAAAATAAATAATAATTTAACGCATCA 2369


RESULT 5
US-09-643-597-306
; Sequence 306, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C11
;  CURRENT APPLICATION NUMBER: US/09/643,597
;  CURRENT FILING DATE:  2000-08-21
;  NUMBER OF SEQ ID NOS:  369
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 306
;   LENGTH: 457
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-643-597-306

  Query Match             7.2%;  Score 34.4;  DB 4;  Length 457;
  Best Local Similarity   54.9%;  Pred. No. 0.87;
  Matches   89;  Conservative    0;  Mismatches   71;  Indels    2;  Gaps    1;

Qy        138 AAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCA--CAGAA 195
              || ||||| |  || ||    | |||  |||| |||||||  |      ||  |
Db          9 AAGGTAATAACTTTTATTATATTAAAGACAATGCAAACGAAAAACAGAATTGAGCAGTGC 68

Qy        196 AAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAA 255
              ||||||  ||  | | |||    | |  || ||| ||  |  | | | | |  ||||| |
Db         69 AAAATTTAAAGGACTGTTTTGTTCTCAAAGTTGCAAGTTTCAAAGCCAAAAGAATTATAT 128

Qy        256 CTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACT 297
              | |   | ||| ||   ||| ||||||  |   | ||| ||
Db        129 GTATCAAATATATAAGTAAAAAAAAGTTAGACTTTCAAGCCT 170


RESULT 6
5428147-1/c
;Patent No. 5428147
;    APPLICANT: BARKER, RICHARD F.;KEMP, JOHN D.
;    TITLE OF INVENTION: OCTOPINE T-DNA PROMOTERS
;    NUMBER OF SEQUENCES: 17
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/91,538
;      FILING DATE: 13-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 869,216
;      FILING DATE: 13-APR-1992
;      APPLICATION NUMBER: 869,216
;      FILING DATE: 13-APR-1992
;      APPLICATION NUMBER: 440,432
;      FILING DATE: 21-NOV-1989
;      APPLICATION NUMBER: 553,786
;      FILING DATE: 19-NOV-1983
;      APPLICATION NUMBER: 741,034
;      FILING DATE: 06-AUG-1991
;      APPLICATION NUMBER: 144,775
;      FILING DATE: 20-JAN-1988
;      APPLICATION NUMBER: 485,614
;      FILING DATE: 15-APR-1983
;      APPLICATION NUMBER: 713,624
;      FILING DATE: 10-JUN-1991
;      APPLICATION NUMBER: 260,574
;      FILING DATE: 21-OCT-1988
;      APPLICATION NUMBER: 848,733
;      FILING DATE: 01-APR-1986
;      APPLICATION NUMBER: 535,354
;      FILING DATE: 26-SEP-1983
;SEQ ID NO:1:
;      LENGTH: 24595
5428147-1

  Query Match             7.1%;  Score 33.8;  DB 6;  Length 24595;
  Best Local Similarity   51.0%;  Pred. No. 4.9;
  Matches   80;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        200 TTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTT 259
              |||   | |  ||      ||  ||| ||||| |  |  | || | |  | |||||||
Db      23880 TTCCCTTCAAGCTGGCTATGATGAACTTGGATAGGGTCTTCCAAAAAAATTTACAACCGA 23821

Qy        260 TGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTC 319
              ||  || | ||| ||||| ||  |  || || |  | ||  |||   |    |  |
Db      23820 ACAAACAAACAAATACAGAAATTTAGACAATAAAAATGAATTACAACCGCATATATCCTG 23761

Qy        320 CACGAACTCGTACCTTATTCCACAAAGGCTTAGACTT 356
              |   | ||  |  |  | || ||||    | ||  ||
Db      23760 CCATCAGTCATCACAAAATCAACAATTAATGAGTTTT 23724


RESULT 7
US-07-638-431-1/c
; Sequence 1, Application US/07638431
; Patent No. 5198535
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard
;    APPLICANT:  Khusmith, Srisin
;    APPLICANT:  Rogers IV, William O.
```

```
;    TITLE OF INVENTION:  Protective malaria sporozoite surface protein
;    TITLE OF INVENTION:  immunogen and gene
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  A. David Spevack
;      STREET:  NMRDC Building 1 T-12 National Naval
;      STREET:  Medical Center
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20814-5044
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/638,431
;      FILING DATE:  19910110
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, Avrom D.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-4033
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4673 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
;      STRAIN:  17X(NL)
;      DEVELOPMENTAL STAGE:  erythrocytic stage
;      TISSUE TYPE:  Blood
;      CELL TYPE:  erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
;      CLONE:  Py10.1111
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  718..3195
;      OTHER INFORMATION:
US-07-638-431-1

  Query Match             7.1%;  Score 33.6;  DB 1;  Length 4673;
  Best Local Similarity   48.0%;  Pred. No. 3.2;
  Matches   96;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy         93 ACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTA 152
              |||  |   |  |||||||||  |  |     | | | ||| ||| ||| | |    ||
Db       4427 ACAAATAAGAATATTATTTCCTTTCCTAATAATATGAACAAGACAAAAATGGATGAAATA 4368

Qy        153 ATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCT 212
                 || |||||| |  ||  |  | ||| |   |   |  ||    |||  | |
Db       4367 TATATACAATACTCAAAATTAATAAATTTACAAAATAAAGAATTAAGCAATAAAGAATTA 4308

Qy        213 TTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATA 272
               |  | || |   |  ||||  || |  |||      || ||   |    |  |||
Db       4307 ATAAATTTATATGCTCCAAACAAAAAATGGACACATGCAACTACAAATAATGAAGAAAAT 4248

Qy        273 TAAAGAAAGTTCTAACATGA 292
               |||  || ||  || ||||
Db       4247 GAAAATAAATTGGAAAATGA 4228


RESULT 8
PCT-US92-00018-1/c
; Sequence 1, Application PC/TUS9200018
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard
;    APPLICANT:  Khusmith, Srisin
;    APPLICANT:  Rogers IV, William O.
;    TITLE OF INVENTION:  Protective malaria sporozoite surface protein
;    TITLE OF INVENTION:  immunogen and gene encoding
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  A. David Spevack
;      STREET:  NMRDC Building 1 T-12 National Naval
;      STREET:  Medical Center
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20814-5044
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00018
;      FILING DATE:  19920103
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, Avram D.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-4033
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4673 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
;      STRAIN:  17X(NL)
;      DEVELOPMENTAL STAGE:  erythrocytic stage
;      TISSUE TYPE:  Blood
;      CELL TYPE:  erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
;      CLONE:  Py10.1111
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  718..3195
;      OTHER INFORMATION:
PCT-US92-00018-1

  Query Match             7.1%;  Score 33.6;  DB 5;  Length 4673;
  Best Local Similarity   48.0%;  Pred. No. 3.2;
  Matches   96;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy         93 ACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTA 152
              |||  |   |  |||||||||  |  |     | | | ||| ||| ||| | |    ||
Db       4427 ACAAATAAGAATATTATTTCCTTTCCTAATAATATGAACAAGACAAAAATGGATGAAATA 4368

Qy        153 ATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCT 212
                 || |||||| |  ||  |  | ||| |   |   |  ||    |||  | |
Db       4367 TATATACAATACTCAAAATTAATAAATTTACAAAATAAAGAATTAAGCAATAAAGAATTA 4308

Qy        213 TTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATA 272
               |  | || |   |  ||||  || |  |||      || ||   |    |  |||
Db       4307 ATAAATTTATATGCTCCAAACAAAAAATGGACACATGCAACTACAAATAATGAAGAAAAT 4248

Qy        273 TAAAGAAAGTTCTAACATGA 292
               |||  || ||  || ||||
```

```
Db       4247 GAAAATAAATTGGAAAATGA 4228


RESULT 9
US-08-955-138-1/c
; Sequence 1, Application US/08955138A
; Patent No. 5977435
; GENERAL INFORMATION:
;  APPLICANT: Lefebvre, Daniel D.
;  APPLICANT: Gellatly, Kevin S.
;  TITLE OF INVENTION: PLANT PHOSPHATASES
;  FILE REFERENCE: PPL97-01
;  CURRENT APPLICATION NUMBER: US/08/955,138A
;  CURRENT FILING DATE: 1997-10-21
;  NUMBER OF SEQ ID NOS: 119
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;    LENGTH: 9412
;    TYPE: DNA
;    ORGANISM: SOLANUM TUBEROSUM
;    FEATURE:
;    NAME/KEY: VARIANT
;    LOCATION: (1)...(17)
;    OTHER INFORMATION: Xaa = Any Amino Acid
US-08-955-138-1

  Query Match             7.1%;  Score 33.6;  DB 2;  Length 9412;
  Best Local Similarity   51.5%;  Pred. No. 4;
  Matches  101;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy        142 TAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATT 201
              ||| |||||||||    |   | | || ||||   | |  |       |||||      |
Db       2535 TAAGCAGCTTAATCTTTTTCTAATTAAGACAATGAAAAATAGTACAGTACAGATTTGGAT 2476

Qy        202 CTAATT-AGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTT 260
              | ||||   ||         ||  | || |||   ||| |  | |  | ||||| || ||
Db       2475 CCAATTCTCTCAACATTATGAGCCAGTGAAAAGGCAATGCACAAAAAATTTACAGCTCTT 2416

Qy        261 GCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCC 320
              ||| ||| |   || | ||| |   |  |  || |  |  |||| |  | |||||||
Db       2415 GCATATATATAGTAGAAAAAATGAAATAAGAAAAAAAAAAACTAAGCTCTAATCTAGTCC 2356

Qy        321 ACGAACTCGTACCTTA 336
              |   | |  | | | |
Db       2355 ATTCAATGCTGCATCA 2340


RESULT 10
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
;  APPLICANT: Berglind Ran Olafsdottir
;  APPLICANT: Jeffrey Gulcher
;  TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
;  FILE REFERENCE: 2345.2001-000
;  CURRENT APPLICATION NUMBER: US/09/426,290
;  CURRENT FILING DATE: 1999-10-25
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;    LENGTH: 168575
;    TYPE: DNA
;    ORGANISM: Homo Sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (21181)...(21403)
;    NAME/KEY: CDS
;    LOCATION: (95252)...(95430)
;    NAME/KEY: CDS
;    LOCATION: (101753)...(101996)
;    NAME/KEY: CDS
;    LOCATION: (110324)...(110439)
;    NAME/KEY: CDS
;    LOCATION: (124058)...(124278)
;    NAME/KEY: CDS
;    LOCATION: (127009)...(127130)
;    NAME/KEY: CDS
;    LOCATION: (128910)...(129139)
US-09-426-290-1

  Query Match             7.1%;  Score 33.6;  DB 4;  Length 168575;
  Best Local Similarity   49.4%;  Pred. No. 11;
  Matches   87;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        204 AATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCA 263
              || ||   ||| | | | | | | | |  || || |    |  ||| || | |   || |
Db      53841 AAATAAATTTTCCTTATTGTATTAGTGAAATGTAACAATTTATTGATCAATATAAATGAA 53782

Qy        264 AATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACG 323
              | ||| |   ||   | ||| ||| |    || |   || ||  |||||    | |
Db      53781 ATAAAATTTAGAATTTATTTCCAACTTTGCTACAATGCATAGCAACGAGACTCCATCTCA 53722

Qy        324 AACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCAT 379
              |   ||    |||||  | |  ||||||  ||   | |    |||||  |   ||
Db      53721 ATAATCTATGAATTTCCAAAGATTCTTAGATTTATAAGATGCAAAGATTCTTAGAT 53666


RESULT 11
US-08-591-629-1/c
; Sequence 1, Application US/08591629
; Patent No. 5993808
;  GENERAL INFORMATION:
;    APPLICANT:  MELCHERS, Leo Sjoerd
;    APPLICANT:  APOTHEKER-DE GROOT, Marion
;    APPLICANT:  BOL, John Ferdinand
;    APPLICANT:  CORNELISSEN, Bernardus Johannes Clemens
;    APPLICANT:  LINTHORST, Hubertus Josephus Maria
;    APPLICANT:  PONSTEIN, Anne Silene
;    APPLICANT:  SELA-BUURLAGE, Marianne Beatrix
;    TITLE OF INVENTION:  Plant chitinases, DNA coding  therefor and
;    TITLE OF INVENTION:  plants containing same
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ladas & Parry
;      STREET:  26 West 61st Street
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023-7604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM PC 4.86 SX 50 Mhz
;      OPERATING SYSTEM:  DOS 6.20
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,629
;      FILING DATE:  15-FEB-96
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/02761
;      FILING DATE:  17-AUG-94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 93202425.0
;      FILING DATE:  17-AUG-93
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MASS, CLIFFORD J.
;      REGISTRATION NUMBER:  30,086
;      REFERENCE/DOCKET NUMBER:  U-010627-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 708-1800
;      TELEFAX:  (212) 246-8959
;      TELEX:  233288
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1253 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nicotiana tabacum
;      STRAIN:  Samsun NN
;      DEVELOPMENTAL STAGE:  TMV-induced
;    IMMEDIATE SOURCE:
;      CLONE:  Cluster-A
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14..1126
;      OTHER INFORMATION:  /partial
US-08-591-629-1

  Query Match             7.0%;  Score 33.2;  DB 2;  Length 1253;
  Best Local Similarity   53.0%;  Pred. No. 2.6;
  Matches   71;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         54 TTTCTGTTGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCC 113
              ||| ||  |  |||||  || || |  ||  || || | | |||||  |||  | || ||
Db       1232 TTTTTGGAGACCACAATTTATTGTTTATTTATTTATGTCATAGTTCACATATCTGTTGCC 1173

Qy        114 ATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAA 173
              | | || ||      || ||||  |  |||||  | |  |  |      ||  | |
Db       1172 TTATTATAAGTAGTATACTTGACACACACAATTACGTAATCCATCACTTCATCTCTTGAG 1113

Qy        174 ACGAGATGACACAT 187
              | || |   | |||
Db       1112 ATGACACTCCCCAT 1099


RESULT 12
US-08-392-625-16/c
; Sequence 16, Application US/08392625
; Patent No. 5837485
;  GENERAL INFORMATION:
;    APPLICANT:  Entian, Karl-Dieter
;    APPLICANT:  G tz, Friedrich
;    APPLICANT:  Schnell, No. 5837485bert
;    APPLICANT:  Augustin, Johannes
;    APPLICANT:  Engelke, Germar
;    APPLICANT:  Rosenstein, Ralf
;    APPLICANT:  Kaletta, Cortina
;    APPLICANT:  Klein, Cora
;    APPLICANT:  Wieland, Bernd
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Jung, G nther
;    APPLICANT:  Kellner, Roland
;    TITLE OF INVENTION:  Biosynthetic Process For The Preparation
;    TITLE OF INVENTION:  Of Chemical Compounds
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/392,625
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/876,791
;      FILING DATE:  30-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.0980002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-392-625-16

  Query Match             7.0%;  Score 33.2;  DB 2;  Length 8700;
  Best Local Similarity   47.6%;  Pred. No. 5;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        128 GCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACAT 187
              | |  | ||||||  ||  |      |   ||    | ||||  |    |  | || |
Db       8306 GAAGTTCACTAAATGAAGAATTAAGTTATTATAGTGTTGAATATGATAATGCAAAAACTT 8247

Qy        188 CCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATT 247
              | ||||  |||  |   |    |  || ||  |  |  |||  |    | ||||||
Db       8246 TTAAAGAATCCATTAAACAAAAAAATATCGAGTTAACTTATAAAATTCCTGAACTACATA 8187

Qy        248 AATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACA 307
              |  |||| ||       |  ||||  |||  || |||    || |||| || |   |||
Db       8186 CAGCACAAATTAAAACTAGCAAATCAAAATTAAATTCACTAATAAAGAGTAATAAAAACG 8127

Qy        308 TGAAGACTAGTCCACGAACTCGTACC 333
              | |||  |      |  ||  || ||
Db       8126 TCAAGTTTGTAAATCCTACCTGTTCC 8101


RESULT 13
US-08-466-961A-16/c
; Sequence 16, Application US/08466961A
; Patent No. 5843709
;  GENERAL INFORMATION:
;    APPLICANT:  Entian, Karl-Dieter
;    APPLICANT:  G tz, Friedrich
;    APPLICANT:  Schnell, No. 5843709bert
;    APPLICANT:  Augustin, Johannes
;    APPLICANT:  Engelke, Germar
;    APPLICANT:  Rosenstein, Ralf
;    APPLICANT:  Kaletta, Cortina
;    APPLICANT:  Klein, Cora
;    APPLICANT:  Wieland, Bernd
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Jung, G nther
;    APPLICANT:  Kellner, Roland
;    TITLE OF INVENTION:  Biosynthetic Process for the Preparation of
;    TITLE OF INVENTION:  Chemical Compounds
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, NW
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,961A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/392,625
;      FILING DATE:  22-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/876,791
;      FILING DATE:  30-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/784,234
;      FILING DATE:  31-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.0980004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-466-961A-16

  Query Match             7.0%;  Score 33.2;  DB 2;  Length 8700;
  Best Local Similarity   47.6%;  Pred. No. 5;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        128 GCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACAT 187
              | |  | ||||||  ||  |       |    ||    | ||||  |    |   | || |
Db       8306 GAAGTTCACTAAATGAAGAATTAAGTTATTATAGTGTTGAATATGATAATGCAAAAACTT 8247

Qy        188 CCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATT 247
              | ||||   |||  |   |    | || ||  | |   ||| |    | ||||||
Db       8246 TTAAAGAATCCATTAAACAAAAAAATATCGAGTTAACTTATAAAATTCCTGAACTACATA 8187

Qy        248 AATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACA 307
              |  |||| ||       |  ||||  |||  || |||    || |||| || |   |||
Db       8186 CAGCACAAATTAAAACTAGCAAATCAAAATTAAATTCACTAATAAAGAGTAATAAAAACG 8127

Qy        308 TGAAGACTAGTCCACGAACTCGTACC 333
              | |||  |      |  ||  || ||
Db       8126 TCAAGTTTGTAAATCCTACCTGTTCC 8101


RESULT 14
US-08-645-193B-18/c
; Sequence 18, Application US/08645193B
; Patent No. 5962253
;  GENERAL INFORMATION:
;    APPLICANT:  Kupke, Thomas
;    APPLICANT:  Gotz, Friedrich
;    APPLICANT:  Kempter, Christoph
;    APPLICANT:  Jung, Gunther
;    TITLE OF INVENTION:  Oxidative Decarboxylation of Peptides
;    TITLE OF INVENTION:  Catalyzed by Flavoprotein EpiD
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/645,193B
;      FILING DATE:  13-MAY-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REGISTRATION NUMBER:  32,893
;      REFERENCE/DOCKET NUMBER:  0652.1540000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
US-08-645-193B-18

  Query Match             7.0%;  Score 33.2;  DB 2;  Length 8700;
  Best Local Similarity   47.6%;  Pred. No. 5;
  Matches   98;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        128 GCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACAT 187
              | |  | ||||||  ||  |       |    ||    | ||||  |    |   | || |
Db       8306 GAAGTTCACTAAATGAAGAATTAAGTTATTATAGTGTTGAATATGATAATGCAAAAACTT 8247

Qy        188 CCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATT 247
              | ||||   |||  |   |    | || ||  | |   ||| |    | ||||||
Db       8246 TTAAAGAATCCATTAAACAAAAAAATATCGAGTTAACTTATAAAATTCCTGAACTACATA 8187

Qy        248 AATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACA 307
              |  |||| ||       |  ||||  |||  || |||    || |||| || |   |||
Db       8186 CAGCACAAATTAAAACTAGCAAATCAAAATTAAATTCACTAATAAAGAGTAATAAAAACG 8127

Qy        308 TGAAGACTAGTCCACGAACTCGTACC 333
              | |||  |      |  ||  || ||
Db       8126 TCAAGTTTGTAAATCCTACCTGTTCC 8101


RESULT 15
US-08-666-405-12
; Sequence 12, Application US/08666405
; Patent No. 5874220
;  GENERAL INFORMATION:
;    APPLICANT:  FACH, Patrick; GUILLOU,
;    APPLICANT:  Jean-Pierre;  POPOFF, Michel
;    TITLE OF INVENTION:  PRIMERS FOR THE
;    TITLE OF INVENTION:  AMPLIFICATION OF GENES CODING FOR THE
;    TITLE OF INVENTION:  ENTEROTOXIN AND THE LECITHINASE OF CLOSTRIDIUM
;    TITLE OF INVENTION:  PERFRINGENS AND THEIR APPLICATION TO THE
;    TITLE OF INVENTION:  DETECTION AND NUMERATION OF THESE BACTERIAE
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN & MUSERLIAN
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/666,405
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP94/04292
;      FILING DATE:  22-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/172,026
;      FILING DATE:  22-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MUSERLIAN, CHARLES A
```

; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 102.164
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: pMRP126
US-08-666-405-12

Query Match 6.9%; Score 33; DB 2; Length 357;
Best Local Similarity 54.5%; Pred. No. 2;
Matches 66; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy        141 GTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAAT 200
              ||| | |  || || |    | | |||||  |||| | | || | | | | | |
Db         43 GTATATGGTCAAAATCCTGATTCTGCTAGACAAATAAGATTATATATAGAAAAAGGACAA 102

Qy        201 TCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTT 260
              |||  ||   |  | ||  |||| |||||  | |||| ||| ||||| |  |||
Db        103 TCTTTCTATAAATATAGAATAAGAATTAAAAACTTTACACCTGCATCAATTAGAGTATTT 162

Qy        261 G 261
              |
Db        163 G 163
```

Search completed: July 20, 2003, 05:20:04
Job time : 46 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 05:19:14 ; Search time 171 Seconds
(without alignments)
5730.558 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq3:*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	48.4	10.2	15649	15	US-10-239-676-103	Sequence 103, App
c	2	46	9.7	6665	15	US-10-239-676-4	Sequence 4, Appli
c	3	45.4	9.6	18997	15	US-10-172-086-18	Sequence 18, Appl
c	4	43.2	9.1	11812	15	US-10-239-676-210	Sequence 210, App
c	5	42.8	9.0	513509	12	US-09-754-853A-4	Sequence 4, Appli
c	6	42.4	8.9	15732	15	US-10-239-676-96	Sequence 96, Appl
c	7	41.8	8.8	4604	15	US-10-239-676-200	Sequence 200, App
c	8	41	8.6	18997	15	US-10-172-086-17	Sequence 17, Appl
c	9	40.8	8.6	6158	15	US-10-239-676-24	Sequence 24, Appl
c	10	40.2	8.5	7195	15	US-10-239-676-30	Sequence 30, Appl
c	11	39.8	8.4	414	11	US-09-960-352-6528	Sequence 6528, Ap
c	12	39.8	8.4	6050	15	US-10-172-086-50	Sequence 50, Appl
	13	39.4	8.3	640681	11	US-09-790-988-1	Sequence 1, Appli
c	14	39	8.2	6167	15	US-10-239-676-41	Sequence 41, Appl
	15	38.8	8.2	864	11	US-09-938-842A-4290	Sequence 4290, Ap
c	16	38.8	8.2	6061	15	US-10-239-676-40	Sequence 40, Appl
c	17	38.8	8.2	6478	15	US-10-239-676-123	Sequence 123, App
c	18	38.4	8.1	3158	15	US-10-078-929-187	Sequence 187, App
c	19	38.4	8.1	6074	15	US-10-172-086-43	Sequence 43, Appl
c	20	38.4	8.1	12405	15	US-10-239-676-35	Sequence 35, Appl
	21	38.2	8.0	2000	11	US-09-887-576-204	Sequence 204, App
c	22	38	8.0	8842	15	US-10-239-676-72	Sequence 72, Appl
c	23	37.4	7.9	7432	15	US-10-172-086-3	Sequence 3, Appli
c	24	37.2	7.8	1347	15	US-10-027-632-212736	Sequence 212736,
c	25	37.2	7.8	7040	15	US-10-172-086-14	Sequence 14, Appl
c	26	37.2	7.8	15732	15	US-10-239-676-95	Sequence 95, Appl
c	27	37	7.8	543	15	US-10-027-632-109062	Sequence 109062,
c	28	37	7.8	543	15	US-10-027-632-109063	Sequence 109063,
c	29	37	7.8	1448	8	US-08-781-986A-709	Sequence 709, App
c	30	37	7.8	5834	15	US-10-239-676-135	Sequence 135, App
c	31	37	7.8	6158	15	US-10-239-676-23	Sequence 23, Appl
c	32	37	7.8	177556	12	US-09-952-213D-6	Sequence 6, Appli
c	33	36.8	7.7	446	11	US-09-960-352-3400	Sequence 3400, Ap
	34	36.8	7.7	624	15	US-10-027-632-214763	Sequence 214763,
	35	36.8	7.7	624	15	US-10-027-632-214764	Sequence 214764,
c	36	36.8	7.7	8996	15	US-10-239-676-211	Sequence 211, App
c	37	36.8	7.7	9539	15	US-10-239-676-52	Sequence 52, Appl
c	38	36.8	7.7	10286	15	US-10-239-676-13	Sequence 13, Appl
c	39	36.8	7.7	14649	15	US-10-239-676-121	Sequence 121, App
c	40	36.6	7.7	5690	15	US-10-239-676-74	Sequence 74, Appl
c	41	36.6	7.7	11036	15	US-10-239-676-118	Sequence 118, App
c	42	36.4	7.7	6621	15	US-10-172-086-9	Sequence 9, Appli
c	43	36.4	7.7	7089	15	US-10-239-676-68	Sequence 68, Appl
c	44	36.4	7.7	7588	15	US-10-239-676-191	Sequence 191, App
c	45	36.2	7.6	12465	15	US-10-239-676-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-10-239-676-103/c
; Sequence 103, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 103
; LENGTH: 15649
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-103

Query Match 10.2%; Score 48.4; DB 15; Length 15649;
Best Local Similarity 46.7%; Pred. No. 0.047;
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 136 CTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAA 195
Db 11798 CTAAACAACCAAACCTACTCTTCCAACTCCCCAAAAATCAATTTTTTCATAAAAAAAACA 11739

```
Qy        196 AAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAA 255
              | ||  ||    |   |    | |  |   |  ||| | |||    | | |||  |  |
Db      11738 ACAAAACTTTAAAACACTAAACTATCTATTCTTAAAATTAAATTAAAAAACTAAAAAATA 11679

Qy        256 CTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACT 315
              | ||  |    || |   || | | ||  | ||       | |  || ||| | || | |
Db      11678 CATTAACTTTCTAATTTATAACACAATTTAACCACATCCTCAACCTCAAACCTCAAAAAT 11619

Qy        316 AGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATC 375
                | |      |  |   || ||  ||||  |    || | | | || | | | ||
Db      11618 TTTTCTTTCCATTTTTTTTTTTTAAACAAAATTTCGCTCTATTCCCCAAACTAAAATACA 11559

Qy        376 CCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTC 435
                |  ||    |  || |  | |||  |   ||| ||||   |   | |||| ||| ||
Db      11558 ATAACGATATCTCAACTCACTACAACCTCCGCCTCTAAAATTCAAACAATTCTCCTACTT 11499

Qy        436 CAGACTCATCTAACTCAAAAACAACACACA 465
              ||  ||| |  |   | ||||| ||| | |
Db      11498 CAACCTCCTAAATAACTAAAACTACAAATA 11469


RESULT 2
US-10-239-676-4/c
; Sequence 4, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 4
;   LENGTH: 6665
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-4

  Query Match             9.7%;  Score 46;  DB 15;  Length 6665;
  Best Local Similarity   51.5%;  Pred. No. 0.13;
  Matches  106;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy         88 ATTTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTA 147
              || | | |  ||  |    || |||||   |   | | || || |  | ||||  |||||
Db       3168 ATATAATAACTCACACCTATAATTCCAACACTCTAAAAAACCAAAATAATAAATCAATTA 3109

Qy        148 GCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATT 207
                  |     | |||  |||   || | | || | | | ||  |    |    | ||||
Db       3108 AAACCAAAAATTCAAAACCAAACTAACCAACATAATAAAACCCCCCCATCTCTACTAAAA 3049

Qy        208 AGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATA 267
              | |    |  | |||  |  | | | |||||||| | |    || ||||    || ||
Db       3048 ATACAAAAATTAAATAAACTTAATAATAAATACCTATAATCTCAACTACTTAAAAAACTA 2989

Qy        268 AAATATAAAGAAAGTTCTAACATGAA 293
              ||||| ||| |    |  ||| | ||
Db       2988 AAATAAAAAAATCACTTAAACCTAAA 2963


RESULT 3
US-10-172-086-18/c
; Sequence 18, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Method and nucleic acids for the differentiation
;  TITLE OF INVENTION:  of prostate tumors
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/172,086
;  CURRENT FILING DATE:  2002-06-13
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 18
;   LENGTH: 18997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-18

  Query Match             9.6%;  Score 45.4;  DB 15;  Length 18997;
  Best Local Similarity   49.0%;  Pred. No. 0.31;
  Matches  121;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy         99 CCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCC 158
              | ||   |  |||  |   | | | |     | | | | ||  || ||   ||     |
Db      11885 CAAAACCTATTTTTAAACCCGTTATAACTAAAATTAAATTAACAAACTAAACTACCTTTC 11826

Qy        159 TAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGT 218
              |      ||||| |||| | || ||  |||   | ||||||  |||||  |        |
Db      11825 TCCTCTCCAATAAAAACTAAATAACTTATCATTAAAAAAAAACCTAATAAAAAAACATAT 11766

Qy        219 GTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGA 278
                | ||       | | || |     |  | || ||| | |   |||||| | ||||  |
Db      11765 TAATAATAAAACCAATAAAAAAAATTAAAATTTCCAAATCTATTAAATAACAAATAATAA 11706

Qy        279 AAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATT 338
              || |   || || || || | ||| || |  || |  | ||     | | |   |||
Db      11705 AACTATAAAAATAAAAACAACTTCAAAAAAAAAAAAAAATTCATAATAATTCTTAATTACC 11646

Qy        339 CCACAAA 345
              || |||
Db      11645 ACAAAAA 11639


RESULT 4
US-10-239-676-210/c
; Sequence 210, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 210
;   LENGTH: 11812
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
```

```
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-210

  Query Match             9.1%;  Score 43.2;  DB 15;  Length 11812;
  Best Local Similarity   46.8%;  Pred. No. 0.92;
  Matches  178;  Conservative    0;  Mismatches  193;  Indels    9;  Gaps    1;

Qy         90 TTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGC 149
              ||    | ||  ||| | |||||| ||   |  | | ||  | |     ||| |  |||
Db       1955 TTTTTATTTAAAATTAATATTTCTATAAAACAAAATAATAAAAAATTAAAATTTTTTAAA 1896

Qy        150 TTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAG 209
              | |||    ||  |   |  ||   |      ||| || |||| | | || | || | |
Db       1895 TAAATTTTTTATCAAATATAACGTTCATTCATACAAATACACAAATATAAATATACTAAT 1836

Qy        210 TCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAA 269
                |      ||    ||  ||| | |||| | |    | ||| |       ||   |||
Db       1835 CAATAATACATATTTTTTTCAAAATAAATAACGAATACATTTATATAACCACCACTCAAA 1776

Qy        270 ATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCG 329
              ||  ||  ||    | |||   |  ||  || |           |  |          ||
Db       1775 TTAATAAATAAAACATTACAAACATCCTTTTTTTT---------TTTTTTTTTTTTTCT 1725

Qy        330 TACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGG 389
              |  || |   ||||  |||| || |   | || | | ||    |  |     | |
Db       1724 TTTTTTTAAAACAAAATCTTACTCTATAACCCAAACTAAAATACAATAACGCAATCTCG 1665

Qy        390 ACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAAC 449
              || |  | ||||||   |||   ||   |    ||||| ||| |  ||  |||    |
Db       1664 ACTCACTACAAATTCCACCTCCTAAATTCACGTCATTCTCCTACCTCAACCTCCCGAATA 1605

Qy        450 TCAAAAACAACACACAACCA 469
                | ||||| ||| ||  |||
Db       1604 ACTAAAACTACAAACGCCCA 1585


RESULT 5
US-09-754-853A-4/c
; Sequence 4, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
;  APPLICANT: Hauge, Brian M.
;  APPLICANT:  Parnell, Laurence D.
;  APPLICANT:  Parsons, Jeremy D.
;  APPLICANT:  Wang, Ming Li
;  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;  TITLE OF INVENTION:  Soybean Cyst Nematode Resistance
;  FILE REFERENCE: 38-10(15810)B
;  CURRENT APPLICATION NUMBER: US/09/754,853A
;  CURRENT FILING DATE:  2001-01-05
;  PRIOR APPLICATION NUMBER: US 60/174,880
;  PRIOR FILING DATE: 2000-01-07
;  NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
;   LENGTH: 513509
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (111805)..(113968),(114684)..(115204)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(513509)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: 318013_region_A3
US-09-754-853A-4

  Query Match             9.0%;  Score 42.8;  DB 12;  Length 513509;
  Best Local Similarity   49.5%;  Pred. No. 7.1;
  Matches  110;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        120 TGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGA 179
              | |||   | |   || ||| |  |     || |  ||||  ||| ||  ||  |   |

Db     222302 TCAGAAGGGAATTCTACAAAATTGGTCCTGCTACTAGCCTACTACTGAAACCACTCCTTA 222243

Qy        180 TGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATA 239
                ||||      |  ||||  |||||| |  | | |     || | |  |||||||   ||
Db     222242 CTACACTGATCCTAAAAACCTTCTAAATCCTGTGTCTTAGTGGCATTTGGAAACAACATT 222183

Qy        240 CCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAG 299
              | |  ||||    | | | || |   | |  |   |||| ||   |  | |  | |
Db     222182 GCAAATTTAAAATAACCCTATGAATCAATAGAAATCAGAAACTTTCTATTTTATTAATGC 222123

Qy        300 TTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCA 341
                |||| ||     || ||  ||    ||    ||||||
Db     222122 AGAAAACATTAACTTCTGTACAATAAAATATAATCAATTCCA 222081


RESULT 6
US-10-239-676-96/c
; Sequence 96, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 96
;   LENGTH: 15732
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-96

  Query Match             8.9%;  Score 42.4;  DB 15;  Length 15732;
  Best Local Similarity   56.2%;  Pred. No. 1.7;
  Matches   99;  Conservative    0;  Mismatches   76;  Indels    1;  Gaps    1;

Qy        139 AAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAA 198
              || || |||  | ||  |  |||||||    ||| |  | || | |   |||   |||||
Db       6621 AAATATTTACATAAAATCTTTAAAACTTTCCACATATAAAATCATATCACCAATAAAAAA 6562

Qy        199 ATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTAC-ATTAATTACAACT 257
              |    | |||   |||   | |   | ||  | | |  |||  | | ||| |||||   |
Db       6561 AAAAAATTTACCTTTTTTCTTTCTTATTTAAATATTTTATATTTTCTATTTATTACCTAT 6502

Qy        258 TTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGA 313
              |||  | | |||| | ||  ||  ||  | | | || ||||  || ||||  || |
Db       6501 TTTTAATAAAAAAAAAAATAAATATTAAACCCTCAATACTATCTCAAACAAAAATA 6446


RESULT 7
US-10-239-676-200/c
; Sequence 200, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
```

```
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 200
;   LENGTH: 4604
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-200

  Query Match             8.8%;  Score 41.8;  DB 15;  Length 4604;
  Best Local Similarity   47.2%;  Pred. No. 1.4;
  Matches  127;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy        206 TTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAA 265
              ||| | | |    | || ||||    |      |||||  | |   ||| ||      | |  |
Db       3107 TTAATATATAAATTTAAAAATAACACTACAAATACTAATACCCATTTCATAAAATCCCTA 3048

Qy        266 TAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAA 325
              |||||  ||||  || |  |||||| || |      |  ||  | ||    | |  |  ||
Db       3047 TAAAACTTAAAACAAATATTAACATAAAAAAACACTTAAAACTAAATCTAACTTAAATAA 2988

Qy        326 CTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTG 385
              |  | |  || | ||  ||    || ||       | ||      |  |   |||  ||
Db       2987 ATAATTCAATAATTCAACAAACACTAAACACCTACCTAAAACCAAACACTACATATACCA 2928

Qy        386 ATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATC 445
               | ||  ||   ||  |       | ||||   |    ||    | |  | |    |
Db       2927 CATAAAACGATATTAAAAAATAAATCACATACAACCAATCTCTAACCTTTTAAAAATTTT 2868

Qy        446 TAACTCAAAAACAACACACAACCAATCAT 474
               |||| ||||| ||||| ||   |||| |
Db       2867 AAACTAAAAAAAAACACGCACATAATCGT 2839


RESULT 8
US-10-172-086-17/c
; Sequence 17, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Method and nucleic acids for the differentiation
;  TITLE OF INVENTION:  of prostate tumors
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/172,086
;  CURRENT FILING DATE:  2002-06-13
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 17
;   LENGTH: 18997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-17

  Query Match             8.6%;  Score 41;  DB 15;  Length 18997;
  Best Local Similarity   49.3%;  Pred. No. 4.3;
  Matches  107;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy         99 CCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCC 158
              ||||||  ||  |  |    |    |    |  |||| || ||| |||||    |||
Db      10826 CCAATACCTACATTAAATACTAAACGTTACAACATAACAAAAATAATTTCAATAAAATAA 10767

Qy        159 TAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGT 218
              |  ||   ||||   |  | |   ||   ||| | | ||| |   || ||     |
Db      10766 TTCAATAAAATAACGAAAACAAACCAATACCAAAAATAAATTAAAAAATAAATAATTAAA 10707

Qy        219 GTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGA 278
              | ||| |  |||| |    | | || |  | |   | |   ||||||| | ||| |
Db      10706 AAAAAAAATAAAAACATTAATTATTCTTTCAAAAAATTCTATAATAATAAAAAAAAAATA 10647

Qy        279 AAGTTCTAACATGAAGACTAGTTCTAACATGAAGACT 315
              || |  ||||   || | ||  | |||    | || |
Db      10646 AAATAATAACTAAAAAATTATCTATAATCAAACGAAT 10610


RESULT 9
US-10-239-676-24/c
; Sequence 24, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 24
;   LENGTH: 6158
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-24

  Query Match             8.6%;  Score 40.8;  DB 15;  Length 6158;
  Best Local Similarity   48.9%;  Pred. No. 2.8;
  Matches  137;  Conservative    0;  Mismatches  142;  Indels    1;  Gaps    1;

Qy        190 ACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAA 249
              | | |||||| |  || ||  |  | | | ||  ||||  ||| | |  |   ||  ||
Db       5481 ATAAAAAAAACTAAAACTAACCAATACATTTAACAATTAAAAAATCATAAATAACCAAAA 5422

Qy        250 TTACAACTTTTGC-AAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACAT 308
              | |||| || | |||||   |||||  |   ||   |    || |  |  | |     |
Db       5421 AAAAAACTATTCCTAAATATTTTATAATTATTATTACTATTATAAAAAAATCTTTCTATT 5362

Qy        309 GAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGA 368
              |  | |  |         |  |   || |   | |||  || |  |  ||  | || | |
Db       5361 TATAATTTTTTTCTTTTTTTTTTTTTTTTAAAATAAAATCTCACTCAATCTCCCAAACTA 5302

Qy        369 GATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCC 428
              | ||    |  ||    |  || |  |||| |   |||    |   |  ||||||
Db       5301 AAATACAATAACGAAATCTCAACTCACTACAAACTCCACCTCCCGAATTCACGCCATTCT 5242

Qy        429 CCTCCTCCAGACTCATCTAACTCAAAAACAACACACAACC 468
              ||| |  ||  |||  | |  | ||||| ||| ||  ||
Db       5241 CCTACCTCAACCTCCCCAATAACTAAAACTACAAACGCCC 5202


RESULT 10
US-10-239-676-30/c
; Sequence 30, Application US/10239676
```

```
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 30
;   LENGTH: 7195
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-30

  Query Match             8.5%;  Score 40.2;  DB 15;  Length 7195;
  Best Local Similarity   46.6%;  Pred. No. 4.3;
  Matches  129;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        140 AGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAA 199
              | |||| |    ||       ||||   || | ||   | || |   ||    |  |||||
Db       1159 AATAATAACAACAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATATAAACAATTATAACTAAAAA 1100

Qy        200 TTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTT 259
                |||     ||||    |  | ||||   ||    |  |  ||  |||| || | ||||
Db       1099 AACTACAATCTCTTAATTTTAATAAATAAAAATAAAATAATATAATTTAACTAAACCTTT 1040

Qy        260 TGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTC 319
                |||||  | | ||||  |  |||| | ||    |  |  ||    ||  | | ||
Db       1039 CCAAAATATATTTTAAAACACTTTCTTAAATCCTAAAAAAATCCTTAATTTAAAATACCT 980

Qy        320 CACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCAT 379
                |   |  |      ||  |   ||   |||    | | |  ||||  | || |    |
Db        979 TAAATAAACTCCTAATAAACACAAATAACTTTCTATATACTAAAATTAACATAAAAATAA 920

Qy        380 GGACTGATGGACACCATCCAAATTATCCCTATAAATA 416
                | | |   | |   |  || ||| |  | |||| |
Db        919 TAAATCAACAATAAATTAAAACTTAACTATTTAAAAA 883


RESULT 11
US-09-960-352-6528/c
; Sequence 6528, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6528
;   LENGTH: 414
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 28-LIB3058-032-Q1-K1-G11
US-09-960-352-6528

  Query Match             8.4%;  Score 39.8;  DB 11;  Length 414;
  Best Local Similarity   48.5%;  Pred. No. 1.4;
  Matches  110;  Conservative    0;  Mismatches  117;  Indels    0;  Gaps    0;

Qy        101 AATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTA 160
              || ||| | |   || | |  | | ||  | |  | |||  ||||      |||     |
Db        350 AAAAATAAATAAAATATAAAAAAAAAAATATAAAAATAATATAATATAAAAAATAAAAA 291

Qy        161 AAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGT 220
              ||| | |    |||  | |  | |  |  | | ||||||    ||  |    |
Db        290 AAAATAAGATAAAAAAAAAAAAAAAACTAAAAAAAAAAAAAAAAAAAAAAATAAAATAAAAAA 231

Qy        221 AGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAA 280
              | ||||| ||| | ||     | ||  |  | ||  |   |||||||  ||||| | |
Db        230 AAAAATTAAAAATTAAACTAAAATATATAAAAAAAAATAAAAAAATAAATAATAAAAATA 171

Qy        281 GTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACT 327
                |  ||    || |  |    ||| || || | || |  |  || |
Db        170 AATAAAAATAAAAAATAAAAAATAAAATAAAAAATAATAAAATAAAT 124


RESULT 12
US-10-172-086-50/c
; Sequence 50, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Method and nucleic acids for the differentiation
;  TITLE OF INVENTION:  of prostate tumors
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/172,086
;  CURRENT FILING DATE:  2002-06-13
;  NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 50
;   LENGTH: 6050
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-50

  Query Match             8.4%;  Score 39.8;  DB 15;  Length 6050;
  Best Local Similarity   46.5%;  Pred. No. 5.1;
  Matches  128;  Conservative    0;  Mismatches  147;  Indels    0;  Gaps    0;

Qy        100 CAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCT 159
              ||| ||   ||  | |   || | |  | |||   ||||||  |  ||   |||   |
Db       4821 CAACAACCTTTCACTTAATATTACAATAACACCAAACTAAACAATATACACTAACTTCAA 4762

Qy        160 AAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTG 219
              |||  | ||  ||||      |    |  ||||| |||||||  | ||  ||         |
Db       4761 AAATATAAAATCAAATCTCCTTCAAATACCACAAAAAAAATACAAAATAAAAAAAATCTA 4702

Qy        220 TAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAA 279
              | |||      ||||   |||   | |  |    |    ||||  |  |||| || ||   |
Db       4701 TCGAATAAAAAAACACTATAAAAAAAAAAATAAATTTTTTTTCAATTTAAATTAAAATTCA 4642

Qy        280 AGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTC 339
              | |   ||  |  |  ||  || | || |  | |        |  ||       |
Db       4641 AATATAAAATTCTACCCTCCTTTTCACTTTTAAATCCCCAAATAAAAAAAACTCAACACA 4582

Qy        340 CACAAAGGCTTAGACTTTCCACAAATCGAGATTAT 374
                | |  | || |||    ||||  |  |  ||||
Db       4581 TAAATAATCATAAACTACACACATTTTAATTTTAT 4547


RESULT 13
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             8.3%;  Score 39.4;  DB 11;  Length 640681;
  Best Local Similarity   52.8%;  Pred. No. 59;
  Matches   85;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        190 ACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAA 249
              | | |||  |||  ||  ||| ||       | |  ||  |||   ||||     |  ||
Db     611989 AAATAAATTATTTGAAAAAGTATTGAGAAAAATATTTTAAAAAAATAATATAAGTAAAAA 612048

Qy        250 TTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATG 309
              | | |   | |   |||||||||| | |   |  ||| | || |||  |  || ||  |
Db     612049 TCATATGCTATTTTAATAAAATAGATAATTACATCTTATATAAAGTATTTTTGTATTAGC 612108

Qy        310 AAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTT 350
              |||| ||    |  |||    |  |||||  | | ||||||
Db     612109 AAGAATATATAAGAAACAAAAATATTATTTTAAAGAGGCTT 612149


RESULT 14
US-10-239-676-41/c
; Sequence 41, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 41
;   LENGTH: 6167
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-41

  Query Match             8.2%;  Score 39;  DB 15;  Length 6167;
  Best Local Similarity   49.7%;  Pred. No. 8.3;
  Matches   99;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy        276 AGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTT 335
              | ||| |   |  |  |  | || | |  |||| ||  | |   ||  ||||| |||||
Db       2123 ATAAAATAAAACTAATATTAATAATACCCACATAAAACCGAACACAATAACTCATACCTA 2064

Qy        336 ATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCA 395
                  | |||    ||| |    | | | |     || ||   ||  |   || || ||||
Db       2063 TAATCCCAACACTTTAAAAAACCGAAACAAACGAATCATAAAATCAAAAAATCGAAACCA 2004

Qy        396 TCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAA 455
              ||| || || | | ||||| ||  |  | | |       | || |   ||  | | |
Db       2003 TCCTAACTAACACGATAAAACCCCATCTCTACTAAAAATACAAAAAAATTAACCGAACGT 1944

Qy        456 ACAACACACAACCAATCAT 474
              |  |   || ||| || ||
Db       1943 AATAATAACCACCTATAAT 1925


RESULT 15
US-09-938-842A-4290
; Sequence 4290, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4290
;   LENGTH: 864
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-4290

  Query Match             8.2%;  Score 38.8;  DB 11;  Length 864;
  Best Local Similarity   64.4%;  Pred. No. 3.6;
  Matches   58;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy        193 GAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTA 252
              ||||||  ||||| ||| ||||   ||||  | |||||   | || | ||| | |  ||
Db         84 GAAAAAGATCTAAATAGGCTTTATCTGTATGACTTGGAGGGTCAAAAACTAAAGTCTTTT 143

Qy        253 CAACTTTTGCAAATAAAATATAAAGAAAGT 282
                ||| ||   |||||||| |  |  ||||
Db        144 AGACTCTTAGTAATAAAATGTTGAATAAGT 173


Search completed: July 20, 2003, 06:21:49
Job time : 174 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 04:50:04 ; Search time 1127 Seconds
(without alignments)
6825.959 Million cell updates/sec

Title: US-09-719-002-1
Perfect score: 475
Sequence: 1 gaattcttattgcgacctga.....acaacacacaaccaatcatg 475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	61.2	12.9	1101	17	CNS0100X	AL098379 Drosophil
	2	59	12.4	1101	17	CNS0039G	AL063921 Drosophil
	3	48.4	10.2	1101	17	CNS0180R	AL108741 Drosophil
c	4	47.4	10.0	1101	17	CNS0039G	AL063921 Drosophil
c	5	45.6	9.6	1043	17	CNS0145P	AL103735 Drosophil
	6	45.2	9.5	1101	17	CNS0176E	AL107648 Drosophil
	7	44.2	9.3	1084	17	CNS00Z5S	AL097258 Drosophil
	8	44	9.3	358	9	AU262156	AU262156 AU262156
	9	44	9.3	638	9	AL513901	AL513901 AL513901
c	10	43.8	9.2	860	17	CNS011EU	AL100176 Drosophil
c	11	43.8	9.2	1101	17	CNS017PB	AL108329 Drosophil
	12	43.6	9.2	939	17	CNS00CNG	AL059400 Drosophil
	13	43.6	9.2	947	17	BH133700	BH133700 ENTNK79TF
	14	43.4	9.1	767	17	CNS00AQX	AL055924 Drosophil
	15	43.4	9.1	919	17	CNS005RL	AL061409 Drosophil
	16	43.2	9.1	1101	17	CNS00L8X	AL068016 Drosophil
	17	43	9.1	301	17	CNS03C4J	AL237340 Tetraodon
	18	43	9.1	848	17	CNS00GWQ	AL072642 Drosophil
c	19	43	9.1	1086	17	CNS00YXK	AL096962 Drosophil
c	20	43	9.1	1101	17	CNS000B8	AL063632 Drosophil
c	21	43	9.1	1101	17	CNS0183Y	AL108856 Drosophil
c	22	42.8	9.0	1201	17	CNS0167M	AL106396 Drosophil
c	23	42.6	9.0	827	17	CNS00EU3	AL069854 Drosophil
c	24	42.6	9.0	887	17	AZ528056	AZ528056 ENTBV59TR
	25	42.6	9.0	1101	17	CNS00EVL	AL069706 Drosophil
	26	42.4	8.9	870	17	CNS06UZZ	AL416517 T7 end of
	27	42.4	8.9	1101	17	CNS007LP	AL067223 Drosophil
	28	42.2	8.9	924	17	CNS07A5L	AL436159 T7 end of
c	29	42.2	8.9	1037	14	BQ648574	BQ648574 AGENCOURT
	30	42.2	8.9	1039	17	CNS006H7	AL064351 Drosophil
c	31	42.2	8.9	1077	17	CNS015CB	AL105269 Drosophil
c	32	42.2	8.9	1101	17	CNS002FG	AL062437 Drosophil
c	33	42	8.8	538	17	AQ523324	AQ523324 HS_5231_B
	34	41.8	8.8	878	17	CNS0187R	AL108993 Drosophil
	35	41.8	8.8	1033	17	CNS0143L	AL103659 Drosophil
	36	41.4	8.7	300	17	BH872616	BH872616 hp07c05.b
	37	41.4	8.7	874	17	CNS007YZ	AL051013 Drosophil
c	38	41.4	8.7	926	17	AZ202638	AZ202638 SP_0059_B
c	39	41.4	8.7	1101	17	CNS00AY1	AL056483 Drosophil
c	40	41.4	8.7	1101	17	CNS00LT2	AL078714 Drosophil
	41	41.2	8.7	512	14	C92998	C92998 C92998 Dict
	42	41.2	8.7	622	17	DR7L8T	AL740919 Danio rer
	43	41.2	8.7	1038	17	CNS01TL7	AL166660 Tetraodon
	44	41.2	8.7	1101	17	CNS00Z6Z	AL097301 Drosophil
	45	41	8.6	360	9	AU269640	AU269640 AU269640

ALIGNMENTS

RESULT 1
CNS0100X/c
LOCUS CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL098379
VERSION AL098379.1 GI:5609990
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers

```
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN03G04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      195 a    108 c    131 g    161 t    506 others
ORIGIN

  Query Match             12.9%;  Score 61.2;  DB 17;  Length 1101;
  Best Local Similarity   20.3%;  Pred. No. 0.0002;
  Matches   68;  Conservative  145;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        137 TAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAA 196
              :| : :: :|  | |: :|:||  :::::|: ::|: :: | :: :|:::|: ||:| :|:
Db        712 YATMTYWTWTYHCMTMSHCHCCACMHMYHCHCYMCHYMMHAWMHMWCWMAHACAMAMHAM 653

Qy        197 AAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAAC 256
              :::   :::    |: : : : |: : : :  ::::: : :|:: |:   || : :|:
Db        652 MMMAMAAMMWCCMMCWCWCYMYCTMAHTHMWHWHMWHWWHTHAHHAAHTMAAACWCHAMA 593

Qy        257 TTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTA 316
              |: : : : ::::  | :  : :: :::|:  |: ::: : : :|||: :| : :|
Db        592 THAHTMCHTHMHMTNTTYHTYTCHAYHTYWHCWCWAHAMYKCYHAMMMACAYAMAMMAHA 533

Qy        317 GTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCC 376
              :::: :|:::  : || :|:::| |::  :: : ::: :  |  :::: : :  |: :
Db        532 MAYMMMTMAMHMHAWWCTNHTWBHAWAMHAHMHAMWHMWAHAWAACHHHMWMAMAMAHWM 473

Qy        377 CATGGACTGATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCC 436
              ::   | : :   |:: :|: :|:   : : : ::|::   :  : |:  :|:|:|:  |
Db        472 MWCAAAAWAWAWAAMWAHAWWHAHTMCWAHAHAMWAHHAYAHAWMTCMCCHCMCHCMCTC 413

Qy        437 AGACTCATCTAACTCAAAAACAACACACAACCAAT 471
              | || ||:: | | |:|| | ||:|:|:|| ::|:
Db        412 ACACNCAYMCACCCCMAAYAAAAMAHAMAAAMHAH 378


RESULT 2
CNS0039G
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES    Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN

  Query Match             12.4%;  Score 59;  DB 17;  Length 1101;
  Best Local Similarity   13.4%;  Pred. No. 0.00066;
  Matches   52;  Conservative  203;  Mismatches  134;  Indels    0;  Gaps    0;

Qy         85 TTGATTTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAA 144
              :: :|:|     :: :: : :|::::|:: : : :| : :  : :|::: : :::  :|:
Db        570 WYYHTYTTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAW 629

Qy        145 TTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTA 204
              :::  :|||::::::| :  ::::::::|:|: | ::  :  :|:  ::   :  : : |
Db        630 WHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCA 689

Qy        205 ATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAA 264
              :  : |:| : | : ::   |::   :: : : :: ::: |:::::     ::::: ::
Db        690 MMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWMYCTACTYHYHHHHHY 749

Qy        265 ATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGA 324
              ::| : |::|:| |:: ::: :  |  |: |    :::::  :: ::  ::: ::: |
Db        750 HWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYHYYMYTCCY 809

Qy        325 ACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACT 384
              :||:  :::::  : |: : :  ::|: : :::  :::  |: | ::|:| | ::  | |
Db        810 MCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWWWHATT 869

Qy        385 GATGGACACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCAT 444
                :  :|: |:|:::::: : :::::  :|::: : |:|::::|  :::: : ::||::
Db        870 WTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHW 929

Qy        445 CTAACTCAAAAACAACACACAACCAATCA 473
              :::::::::::::|  :  :  :::: |:
Db        930 MWHMHWHWHHWWMATWMTTMTTMMMMCCM 958


RESULT 3
CNS0180R
LOCUS       CNS0180R                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108741
VERSION     AL108741.1  GI:5629045
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES    Location/Qualifiers
```

```
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN37D08"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      329 a    205 c    116 g    304 t    147 others
ORIGIN

  Query Match             10.2%;  Score 48.4;  DB 17;  Length 1101;
  Best Local Similarity   34.2%;  Pred. No. 0.21;
  Matches  104;  Conservative   55;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        165 TCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAA 224
              | |||: || : :    |   |:  | | |||||| :    | | |||||   | ::
Db        717 TTAATWAAATMAWTGNTATTTAWATATAAAAAAAAMWWATGTCATTCTTTYAWTNWWWTT 776

Qy        225 ATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTC 284
              ||   :|| : :  |      |||  ||  || || |  | |:||: | :||| :|:  |
Db        777 ATAAAWAAAWMHCMAAAATAATTTTTTCTAAATTCTTAATAWAAHYTTWAAAAWAMHATT 836

Qy        285 TAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAA 344
              |    | :|  :||   :| ::| :| |  :    |  ||:    :    |:  || |:
Db        837 TTYTCTTWAATMTAAAAWHATMWTTWAAAAAWAAAAAAAAAAMAATMWAAAAAHMTCAAAM 896

Qy        345 AGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAATTA 404
              |   |::   | |:    ||: |  |: || | :   |  :|  |: | :|  :||::|
Db        897 AAAATYMTCTTCTMTCTCTATYCAYCTYTTCTCTYTATCCCMTAWAYCCTMTAAMAAHHA 956

Qy        405 TCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACACAC 464
                | ||| | :  :::|:   : : :  |  :::| |||  ||:: ||| || | :
Db        957 ATAAAAAAAAAAAAMAWMYMAWCTTMTYAHCTCTMWMYCYTCTCTCTMWCAAAAAAAAAMA 1016

Qy        465 AACC 468
              :::|
Db       1017 MMMC 1020


RESULT 4
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES    Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN

  Query Match             10.0%;  Score 47.4;  DB 17;  Length 1101;
  Best Local Similarity   26.7%;  Pred. No. 0.36;
  Matches   92;  Conservative  104;  Mismatches  148;  Indels    1;  Gaps    1;

Qy         92 GACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTT 151
              :: | ::  :|:::|: |::|  :|:| ::: :::::: | : :: ||:| ::::|:: ::
Db        682 RDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAWD 623

Qy        152 AATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTC 211
              :|:    |:|:: : :| : : :  ::|::: | |    : ::::|:||    ::|| |:
Db        622 WAKWDWKTRADRWDRWAADTWTD-ARKADRDWAKARAWRARRDRARAARADRRWTTKGKT 564

Qy        212 TTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAAT 271
              ||    |  |:||::   |:| |" | :   :   :|  ||     |:||  |:::||::|
Db        563 TTATWTTWAARAAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWT 504

Qy        272 ATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTA 331
              ||:||  || :   || |: |    |  || | : :| || :  : :         |   |
Db        503 ATWAAWTAAAWAAAAAAAWAATTTTTTTTTTTTWAWTAAAWWTAWTWTTTTWTTTTWWAA 444

Qy        332 CCTTATTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGAC 391
                || |:   ::|   |   ||| | |||   : ::: :   ||: |   :   |
Db        443 TTTTTTWTWTTWWAATTATTTTWTTTWTAAAAAAAAWWWWWTWTATATKCCCCCCYCCCC 384

Qy        392 ACCATCCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCC 436
              |:  :|   |  ::|||     : ||: ::: ::::::|:| ||
Db        383 CCYCCYCCCCTCCYYCCTCCYTCYCCCYYYYYYYYYYYYYTYCCCC 339


RESULT 5
CNS0145P/c
LOCUS       CNS0145P                1043 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103735
VERSION     AL103735.1  GI:5615346
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES    Location/Qualifiers
     source          1. .1043
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN11G11"
```

```
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT       277 a     96 c    121 g    382 t    167 others
ORIGIN

  Query Match             9.6%;  Score 45.6;  DB 17;  Length 1043;
  Best Local Similarity   36.4%;  Pred. No. 0.96;
  Matches   94;  Conservative   41;  Mismatches  123;  Indels    0;  Gaps    0;

Qy         51 AAATTTCTGTTGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATT 110
              | ||:|  :  :   : || | |    :  :|  :: :::| : : ::  : |  :| ||
Db        840 ATATWTARRGGRAADDAAAAAAAAATTWTTWTWWWWWWWWTTWWWWWWWWTTTTWTTTT 781

Qy        111 TCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATA 170
              |  |    |  | | ||  | |  |  |||  || ||  |:||     ||||  : ||:|
Db        780 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANTATNTWAAATATATAAATTWAAAWA 721

Qy        171 CAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGA 230
               |:|    |  | | ||  | |  |   ||: :|: :: |  |       | ||: :  |
Db        720 TAWATATAAATATAWATAAAAAATATTTATWAWAWAWWATAATATAWAWAAAAAWAWAWA 661

Qy        231 AACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACAT 290
              :|   :|||  |  :::|: || |: : ::  ||| ||:|||||:| |   || :|
Db        660 WAAAAWATATATTTWWWAWATAAAWAWAWWWAAAAAAAWATATAWATATTWTTTWATATA 601

Qy        291 GAAGACTAGTTCTAACAT 308
               || | ||  | ||  ||
Db        600 TAAAAATATATATAKTAT 583


RESULT 6
CNS0176E
LOCUS       CNS0176E                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN17D02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL107648
VERSION     AL107648.1  GI:5627952
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN17D02"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT       337 a     143 c    131 g    350 t    140 others
ORIGIN

  Query Match             9.5%;  Score 45.2;  DB 17;  Length 1101;
  Best Local Similarity   32.7%;  Pred. No. 1.2;
  Matches  114;  Conservative   63;  Mismatches  172;  Indels    0;  Gaps    0;

Qy         65 CACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCATGAGA 124
              :::::::: ||  |  :|  :|  ||:      |: :|:|  | | |||: |       | |
Db        753 HHHHMMVWTCTTTTTTWTTTWTTTTTWTTTTTTWWMAWTTAMTTTTTYGAAAAACAAAAA 812

Qy        125 GAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACA 184
              ||  | |  |  |||     |   |||     ||:| :  :| ||| | |:  :|
Db        813 AAAAAAAAAAAWAAAAMMCCRAAAAAAAATAAARAAAMAAWTTTWAAAAAAAAAAWCCMA 872

Qy        185 CATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTAC 244
              : | |  | ||||||  |||: |.        :  :  ||    |||    | | ::
Db        873 MTTTCTTAAAAAAAAAAATAAWAAAAAAAAAAAWAAWTYAAAATAAAAAACTAAAAHWTA 932

Qy        245 ATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTA 304
              : | |  | ||        ||| |||| | ||| |||     ||::|  | |    |||
Db        933 WCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAHHTNTAAAWCYCCTCTC 992

Qy        305 ACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTCCACAAA 364
              | | :| | |   :::|  :| : : :|||  ::| ::|  |  | |    : :::::
Db        993 CCCTTWAAATTYCCYYHCCYMCMMAWCYCTTAAAMMAAMMACCCAAAAAAAAAAMAMMMMM 1052

Qy        365 TCGAGATTATCCCATGGACTGATGGACACCATCCAAATTATCCCTATAA 413
              : : |  : ||| :   |  :   |:| :: |::::  : :  ||::|
Db       1053 MMMMAAMMMMCCCYYMMCCCCMCCCAMAAMMACMMMHAAMCMTATAWWA 1101


RESULT 7
CNS00Z5S
LOCUS       CNS00Z5S                1084 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN01E02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097258
VERSION     AL097258.1  GI:5608869
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN01E02"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT       453 a     186 c    117 g    173 t    155 others
ORIGIN

  Query Match             9.3%;  Score 44.2;  DB 17;  Length 1084;
  Best Local Similarity   35.2%;  Pred. No. 2.1;
  Matches  127;  Conservative   59;  Mismatches  173;  Indels    2;  Gaps    1;

Qy         97 TTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCC 156
              ||  || ||  |  :|  | | |  | | ||  | |  | |    |||  |  || :|
Db        724 TTTTAAAAAAAAAMYCTTTTTWAAAAAAWAAWAAWAAAAWTKTTTTAAAAAATTTTWTTT 783
```

```
Qy        157 CCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGC 216
Db        784 TAAAAAAAAAATTWAAAACTTTAAAAAAAAAAAAAAAAAAAAAAAATTTTTTAHCCAATTTA 843

Qy        217 GTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAA 276
Db        844 AWTWAAAAAWWTTGWTYTTTAAAAAWAAAATTTTTHTAAWTTTAAAAATTAAAWAAAAAA 903

Qy        277 GAAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTA 336
Db        904 WWWWWWWWAAWAWWAAAMAWAAAAAAAAHCYCCAAAAAAAAMCTTAAAAAAA--AACCTT 961

Qy        337 TTCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCAT 396
Db        962 TWMMAMWAAAAYWHAMHAMAAAAMYMMMAAMCCMYCTYYAAAAMACTTWAMAMMAAAMAA 1021

Qy        397 CCAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAA 456
Db       1022 ATAAHAAMMCMCTAWAAAAMMMMMCCCAAAMMMAACAWWCTACMMAMCCMMCCCCMAMMC 1081

Qy        457 C 457
Db       1082 C 1082
```

RESULT 8
AU262156
LOCUS AU262156 358 bp mRNA linear EST 09-MAY-2002
DEFINITION AU262156 VS Dictyostelium discoideum cDNA clone VSB644 3', mRNA sequence.
ACCESSION AU262156
VERSION AU262156.1 GI:20518703
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 358)
AUTHORS Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M., Takeuchi,I., Kohara,Y. and Tanaka,Y.
TITLE Population analysis of cDNAs from unicellular and multicellular stages of Dictyostelium discoideum
JOURNAL Unpublished (2002)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.
FEATURES Location/Qualifiers
source 1. .358
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="VSB644"
/clone_lib="VS"
/sex="mat A"
/dev_stage="vegetative"
BASE COUNT 167 a 37 c 56 g 98 t
ORIGIN

Query Match 9.3%; Score 44; DB 9; Length 358;
Best Local Similarity 52.1%; Pred. No. 2.3;
Matches 98; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy         92 GACAGTTCCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTT 151
Db        168 GTCAATTTTTAGAACATGTTCAATGTATTGATAAATGTGTTGGACCAAAATTATTTGCTC 227

Qy        152 AATCCCCTAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAATTAGTC 211
Db        228 AATTAAAGTAAAATATATGGCGAATATGTGAAAAATAGTAATAAAAAATAGTAATATAAA 287

Qy        212 TTTGCGTGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAAT 271
Db        288 AAAAAAAAAAAAAAATTAATAGAGAAAATCCATTATAATAAAAATTTAAAAAAAAAAAT 347

Qy        272 ATAAAGAA 279
Db        348 ATTTAAAA 355
```

RESULT 9
AL513901
LOCUS AL513901 638 bp mRNA linear EST 13-FEB-2001
DEFINITION AL513901 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA006ZG01 3 prime, mRNA sequence.
ACCESSION AL513901
VERSION AL513901.1 GI:12777395
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 638)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA006ZG01"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1).301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 347 a 58 c 15 g 141 t 77 others
ORIGIN

Query Match 9.3%; Score 44; DB 9; Length 638;
Best Local Similarity 39.5%; Pred. No. 2.3;
Matches 98; Conservative 31; Mismatches 119; Indels 0; Gaps 0;

```
Qy        119 ATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAG 178
Db        174 ATAAMMAAAAAAAAWAAAAAAAAAAAAAAAAAATWATTACMAWAAAAAAAAAAAAAAAAAG 233

Qy        179 ATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAAT 238
Db        234 GGGAAAAWATTWTAAWAAWAWTAWAAAWATATTTAAATAAWWWAAAAAAAAAAAAWAAAA 293

Qy        239 ACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTA 298
Db        294 AAATAAAAAAAAAAAAAAAGTTTTTTWAAATTCYMMAAAAAAAAAARTAWMCCTBMAAAMAA 353

Qy        299 GTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGACTTTC 358
Db        354 AAAAAAAAAAAAAMMMMAAACCCCCCCCTTTTTAMCCAMCCCMATAAHMMCHTWMCCCCCA 413

Qy        359 CACAAATC 366
Db        414 AAAAAACC 421
```

```
RESULT 10
CNS011EU/c
LOCUS       CNS011EU                860 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN06H18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL100176
VERSION     AL100176.1  GI:5611787
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .860
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN06H18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      202 a    106 c    116 g    306 t    130 others
ORIGIN

  Query Match             9.2%;  Score 43.8;  DB 17;  Length 860;
  Best Local Similarity   41.2%;  Pred. No. 2.6;
  Matches  131;  Conservative   27;  Mismatches  157;  Indels    3;  Gaps    1;

Qy         54 TTTCTGTTGCGCACAACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCC 113
              ||  ||   |||:    ||:|: | | || |   || ||   ||   ||  |  | |
Db        319 TTNWTGACNCGCWTTCNATRCWCGCCGTTNCNATATATGCGCGT---AAATAAGANTGAA 263

Qy        114 ATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAA 173
              |||  |  | || | :|:||:  ||||  |  |   ||  || | |:: |:  ||:
Db        262 ATGAAAAAATAGCAAMAMATRGAAAAAGAATCTCCTTTSTTCTCATWWCAMAGAAWCTTC 203

Qy        174 ACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAAC 233
              :  : |  | | ||  | | | ||||      |   | |:      :    |::  |||
Db        202 WWMMAAAAAAAAATAAAGAAATAAAAAAAATCTASATTTKACAAATKKCTTAWWWTAAAA 143

Qy        234 TGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAA 293
                ||:|   | | || | | || ::::  || | | ||| ||  || ||| ||  | ||
Db        142 AMAAKAATAAAAATATTAAAAAAWWKKNAAATTTATATAAAATWAATGTTTAAAATTAAA 83

Qy        294 GACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCACAAAGGCTTAGA 353
                 ||||| ||| |       | ||    :||:|   | : : |   | | :   | ||
Db         82 AGATAGTTATAATAATTTCCTTTGTTTTAKAAMTTTAAWHAWTTAATATATWTATTAAGT 23

Qy        354 CTTTCCACAAATCGAGAT 371
              | ||  |   |:   |||
Db         22 CATTTAATTTAWTATGAT 5


RESULT 11
CNS017PB/c
LOCUS       CNS017PB                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37B03 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108329
VERSION     AL108329.1  GI:5628633
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN37B03"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
BASE COUNT      252 a    145 c    148 g    371 t    185 others
ORIGIN

  Query Match             9.2%;  Score 43.8;  DB 17;  Length 1101;
  Best Local Similarity   43.3%;  Pred. No. 2.6;
  Matches  117;  Conservative   14;  Mismatches  139;  Indels    0;  Gaps    0;

Qy         69 ACATACTGGTCCTTGCTTGATTTGACAGTTCCAATAATTATTTCCATGTCATGAGAGAAG 128
              ||| | | |   | ||   | |  | | |   ||||| || | | ||   |:   | ||
Db        299 ACAGATTAGCAGTGGCACAANTAAAAAATAGAAATAANTAATACAATARAAWATAAAAAA 240

Qy        129 CACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACAAACGAGATGACACATC 188
              :|||  |  |||  ||  |   |||      ||||   || | :|  || |  | | |:
Db        239 MACAAAAGAAAAAAAAAAAATATAAAWAATGAAAAAAAAAAAGWATAGAAAATAAAAAWA 180

Qy        189 CACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAACTGAATACCTACATTA 248
              : | ||:|||     || :|       :    | ||| :  |||   || |   | | |
Db        179 AWAAAAAMAAAAAAAAAGWAAAAAAAAMAAAAAAAAAAWAWAAAAAAAAAAAAAAAAAA 120

Qy        249 ATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGAAGACTAGTTCTAACAT 308
              |  | ||      :|: |||| | ||| |||     |: |  || :  | |  |||:|
Db        119 AAAAAAAAAAAAAAAMAMAAAAAAAAAAAAAAAAAAAAAWAAATAAAWACAATATTAAMAC 60

Qy        309 GAAGACTAGTCCACGAACTCGTACCTTATT 338
              |  |    | || |   ||     |||||
Db         59 CATCATCTNTTCAAGTTATCTGTTTTTATT 30


RESULT 12
CNS00CNG
LOCUS       CNS00CNG                939 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL059400
VERSION     AL059400.1  GI:4946964
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR26H16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT       71 a    349 c    104 g    180 t    235 others
ORIGIN

  Query Match             9.2%;  Score 43.6;  DB 17;  Length 939;
  Best Local Similarity   6.5%;  Pred. No. 2.9;
  Matches   12;  Conservative  118;  Mismatches   56;  Indels    0;  Gaps    0;

Qy        283 TCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTATTCCAC 342
              || ::::  :: :: :     ::::  :: ::|: :::::  :::  |::| |:| ::::
Db        383 TCCMMMMMMMMMMMMMMMTMHMMMMMMMMMMTMMHMMMMTTMMHAMTMMCMTMTMMMMM 442

Qy        343 AAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCATCCAAAT 402
              :::  : |: ::   ::::::::|: : :  :|::::   : : ::  :::::::  :::
Db        443 MMMTTMMTMMMMMMMMMMMMMTMMMMMMMMTMMMMMMMTHMMHMMMMMMHHMTMMMMM 502

Qy        403 TATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAACAACAC 462
              :: :::: :::: :::   :::| :::: :: ::  :: :: : ::: ::::::::::|
Db        503 HMMMMMHTHHMMMMMMMMTMMMTMMMMMMMMMMTMMMMHMMMMMMMMMMHHMMMMMMC 562

Qy        463 ACAACC 468
              ::::::
Db        563 MMMMMM 568


RESULT 13
BH133700
LOCUS       BH133700                947 bp    DNA     linear   GSS 07-AUG-2001
DEFINITION  ENTNK79TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH133700
VERSION     BH133700.1  GI:15092761
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence stop: 770.
FEATURES             Location/Qualifiers
     source          1. .947
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      447 a     80 c    165 g    255 t
ORIGIN

  Query Match             9.2%;  Score 43.6;  DB 17;  Length 947;
  Best Local Similarity   52.2%;  Pred. No. 2.9;
  Matches   97;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        113 CATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACA 172
              ||| |  |     |  || ||   |||| ||| ||  |    | | ||| |   |||| |
Db        623 CATATAGTACACTACACATATACATAAAATAAATATATACTACACATAATATAAAATATA 682

Qy        173 AACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAA 232
              ||   ||   | ||  | | | || ||    || ||   |  |   || ||||   ||
Db        683 TACATAATATAAAATATATACATAATATAAAGATAAGGGATACCTAATATAAATATAAAT 742

Qy        233 CTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGA 292
              | ||||  ||||| |  || |  ||  | | | | || |||  | |||   | | | ||
Db        743 ATAAATATATACATAATATAAAGATTAGGTATAGAGAAAATACGGGAAGAGATTATAAGA 802

Qy        293 AGACTA 298
              ||  ||
Db        803 AGTATA 808


RESULT 14
CNS00AQX
LOCUS       CNS00AQX                767 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR21A20 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL055924
VERSION     AL055924.1  GI:4936693
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
```

```
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see  http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osoegawa and
          Aaron Mammoser in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw sp, the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR21A20"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      312 a    105 c     65 g    164 t    121 others
ORIGIN

  Query Match             9.1%;  Score 43.4;  DB 17;  Length 767;
  Best Local Similarity   36.2%;  Pred. No. 3.2;
  Matches   77;  Conservative   40;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        113 CATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCCTAAAACTCAATACA 172
              :|  | |: : | |:  | |  | |||| ::|  :  |  : | | || :| | ||:||:
Db        135 MAAATAAWMMWAAAWAWAAAMTAATAAAAWWAAAWTTTATWMCACATACMATTAAAWACM 194

Qy        173 AACGAGATGACACATCCACAGAAAAAATTCTAATTAGTCTTTGCGTGTAGAAATTGGAAA 232
              |   | :| |    |  :||  :| :: :  ||  :  |: |        ::||:  |||
Db        195 ATAAATWTTAATATTTTWCAATMACMWAWAMAAAAWAACWATAWTAMMTCWWATWCWAAA 254

Qy        233 CTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAGAAAGTTCTAACATGA 292
                || |   ::|  : ||:|  |||:   ||| |:|| |:||| |||  :  :: :| |
Db        255 AAAAAAAAAMMMAAMMTTTWCTTCTTWATWAAAAAMAAAAWAAAAAAAAAWAAMWWMTAA 314

Qy        293 AGACTAGTTCTAACATGAAGACTAGTCCACGAA 325
              : :: : | |: | |  || |  : | :|  ||
Db        315 WTWMAMTTACWCAAAAAAAAAAAMWTWMAAMAA 347


RESULT 15
CNS005RL
LOCUS       CNS005RL                 919 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR12F23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061409
VERSION     AL061409.1  GI:4943512
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 919)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .919
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR12F23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      279 a    124 c     93 g    197 t    226 others
ORIGIN

  Query Match             9.1%;  Score 43.4;  DB 17;  Length 919;
  Best Local Similarity   22.8%;  Pred. No. 3.2;
  Matches   85;  Conservative  123;  Mismatches  163;  Indels    1;  Gaps    1;

Qy         99 CCAATAATTATTTCCATGTCATGAGAGAAGCACATGACTAAAGTAATTAGCTTAATCCCC 158
              | || ||:|    | :| :| |      |  | | |  | | : || ::|  ::::: ::
Db        521 CAAAAAAWTCAAACAWTAWCCTTTTCAATTCCCTTTCCAATWCTATWWAWWWWHHHAHHA 580

Qy        159 TAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATTCTAA-TTAGTCTTTGCG 217
              ::::: ::: ::: :: : :  ||::: |::|: : :     | || ||    ::| |
Db        581 MMHMMMCMMMAMMMCMMCMMMMCACMMMMCMMCMMMCMCNCCACNAATTTTAAAWWTACT 640

Qy        218 TGTAGAAATTGGAAACTGAATACCTACATTAATTACAACTTTTGCAAATAAAATATAAAG 277
              | :| | ||   |||| ||:   || |  || :| ||        ||: |::  |::||
Db        641 TMHACATATAAAAAACAAAAWTWTTAAAAAAAAWAAAAAAAAAAAAAWAAWWTNAHWAAM 700

Qy        278 AAAGTTCTAACATGAAGACTAGTTCTAACATGAAGACTAGTCCACGAACTCGTACCTTAT 337
              :::  :: ::::  | :  | : | ::  :: :||  :  |:   ::  |: : :::
Db        701 WMWMAHMMMMMMAWATWMTMATCWATTWMMAMMMMCTTAHAWAYATTHYWTTWTMAWMH 760

Qy        338 TCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGATGGACACCATC 397
              :   :||  :: :   ||||::: :: : |    || |  :  |:  |   : : :|
Db        761 WAWTTWAAWAMWCMMTATTTCMMMCMMCMHAWCACATWCWTHATAMAWAAWTWAWAHAAW 820

Qy        398 CAAATTATCCCTATAAATACCTGCCCATTCCCCTCCTCCAGACTCATCTAACTCAAAAAC 457
              :|:   | : : | ||:|  || ::::: :: : : |:    :|:  : : : :::::
Db        821 MAWCCMACMAMMACAAMTCACTMHMMWHCHMTMMMTTMTYMTMTHTMMMMCMMMMMMMW 880

Qy        458 AACACACAACCA 469
              :: :  ::::::
Db        881 MWTMWTHHHMMW 892


Search completed: July 20, 2003, 06:08:28
Job time : 1131 secs
```